

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2002, 18:07:04 ; Search time 67 Seconds
(without alignments)
421.629 Million cell updates/sec

Title: US-09-817-198A-2

Perfect score: 1105

Sequence: 1 MAKQYDLVRLLLICDGVG.....LEEEGKPEGPANSSKTCWC 212

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_l01002.*
1: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/cgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/cgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/cgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/cgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1105	100.0	401	AAU17136	Novel signal trans
2	1092	98.8	218	AAAB41604	Human OREX ORF1368
3	832	75.3	188	AAU17555	Novel signal trans
4	546.5	49.5	204	ABR70670	Drosophila melanog
5	545	49.3	213	ABR11916	Human rab8 homolog
6	540	48.9	221	ABP41333	Human ovarian anti
7	532	48.1	246	AAAB58196	Lung cancer associ
8	530.5	48.0	200	AAAB19165	Amino acid sequenc
9	530.5	48.0	200	AAAB09979	Human Rab10 protei
10	530.5	48.0	200	AAAB95340	Human protein sequ

11	530.5	48.0	218	21	AAAB56993	Human prostate can
12	529.5	47.9	218	22	ABG07266	Novel human diagno
13	528.5	47.8	207	22	ABB71647	Drosophila melanog
14	528.5	47.8	207	22	ABG67154	Amino acid sequenc
15	528.5	47.8	207	22	AAAB94628	Human protein sequ
16	526	47.6	201	21	AAAB09982	Canine Rab10 protei
17	523	47.3	199	21	AAAB09980	Human Rab10 protei
18	523	47.3	199	21	AAAB09981	Human Rab10 protei
19	519	47.0	216	21	AAAG08688	Arabidopsis thalia
20	519	47.0	216	21	AAAG53945	Arabidopsis thalia
21	519	47.0	253	21	AAAG53944	Arabidopsis thalia
22	519	47.0	254	21	AAAG08687	Arabidopsis thalia
23	513.5	46.5	215	21	AAAG35215	Zea mays protein f
24	510	46.2	218	21	AAAG19220	Arabidopsis thalia
25	507	45.9	216	21	AAAG08006	Arabidopsis thalia
26	506.5	45.8	209	22	ABG23365	Novel human diagno
27	501	45.3	224	21	ABG47826	Arabidopsis thalia
28	501	45.3	234	21	AAAG47825	Arabidopsis thalia
29	501	45.3	335	22	ABG23366	Novel human diagno
30	478.5	43.3	201	20	AAAY00919	Human Rab protein,
31	478.5	43.3	201	22	AAU28024	Novel human secret
32	478.5	43.3	221	22	AAO13525	Human polypeptide
33	478.5	43.3	224	21	AAAB58758	Breast and ovarian
34	476	43.1	203	21	AAAG30499	Arabidopsis thalia
35	476	43.1	258	21	AAAG30498	Arabidopsis thalia
36	468	42.4	202	21	AAAG07763	Arabidopsis thalia
37	467	42.3	205	21	AAAB34843	Gene 44 human secr
38	467	42.3	205	21	AAAB34844	Human secreted pro
39	466	42.2	202	21	AAAG10858	Arabidopsis thalia
40	462	41.8	205	22	ABBS9808	Drosophila melanog
41	457	41.4	218	21	AAAG30710	Arabidopsis thalia
42	457	41.4	221	21	AAAG30709	Arabidopsis thalia
43	457	41.4	254	21	AAAG45323	Arabidopsis thalia
44	456.5	41.3	227	21	AAAB19163	Amino acid sequenc
45	456.5	41.3	227	22	AAAB39600	Human polypeptide

ALIGNMENTS

RESULT 1
AAU17136
ID AAU17136 standard; Protein; 401 AA.
XX
AC AAU17136;
XX
DT 07-NOV-2001 (first entry)
DE
DE
XX
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antineoplastic; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disease; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; AIDS;
KW reproductive system; gastrointestinal; liver disorder; AIDS;
KW acquired immune deficiency syndrome.
OS
OS Homo sapiens.
XX
XX WO200154733-A1.
PD
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01312.
XX
XX 31-JAN-2000; 2000US-0179065.
PR
PR 04-FEB-2000; 2000US-0180628.
PR
PR 24-FEB-2000; 2000US-0184664.
PR
PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 03-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465460/50.

N-PSDB; AAS27053.

Novel polypeptides useful for diagnosing, treating, preventing and/or
 prognosing disorders related to the proteins, including cancers, immune
 disorders and neuronal disorders -

Claim 1; SEQ ID No 701; 880pp; English.

XX

CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.

Query Match 100.0%; Score 1105; DB 22; Length 401;
Best Local Similarity 100.0%; Pred. No. 2.1e-108;
Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAKQYDVLFRLLLLIGDSGVGKTCCLLCRFTDNEFHSHISTIGVDFKMKTIEVDGKVRQ 60
Db 27 MAKQYDVLFRLLLLIGDSGVGKTCCLLCRFTDNEFHSHISTIGVDFKMKTIEVDGKVRQ 86
QY 61 IWDTAGOERYQITTKQYRRAGQIFLVYDISSERSYOHIMKMWSDVDEYAPGVQKILIG 120
Db 87 IWDTAGOERYQITTKQYRRAGQIFLVYDISSERSYOHIMKMWSDVDEYAPGVQKILIG 146
QY 121 NKADEBQKRVGREQOQQLAKYGMDFYETSACTNLNLIKESFTRLTVELVLAHRKEGL 180
Db 147 NKADEBQKRVGREQOQQLAKYGMDFYETSACTNLNLIKESFTRLTVELVLAHRKEGL 206
QY 181 RMRASNELALAELEEEGKPEGPNSSKTCWC 212
Db 207 RMRASNELALAELEEEGKPEGPNSSKTCWC 238

RESULT 2
AAB41604
ID AAB41604 standard; Protein; 218 AA.
XX AAB41604;
XX AAB41604;
DT 08-FEB-2001 (first entry)
DE Human ORF1368 polypeptide sequence SEQ ID NO:2736.
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

XX Homo sapiens.
XX ID
XX WO200058473-A2.

XX 05-OCT-2000.
PD
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
PA
XX Shinkets RA, Leach M;
PI
XX WPI; 2000-602362/57.
DR N-PSDB; AAC75813.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 1979-1980; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 218 AA;
Query Match 98.8%; Score 1092; DB 21; Length 218;
Best Local Similarity 97.2%; Pred. No. 2e-107;
Matches 212; Conservative 0; Mismatches 0; Indels 6; Gaps 1;
QY 1 MAKQYDVLFRLLLLIGDSGVGKTCCLLCRFTDNEFHSHISTIGVDFKMKTIEVDGKVRQ 60
Db 1 MAKQYDVLFRLLLLIGDSGVGKTCCLLCRFTDNEFHSHISTIGVDFKMKTIEVDGKVRQ 60
QY 61 IWDTAGOERYQITTKQYRRAGQIFLVYDISSERSYOHIMKMWSDVDEYAPGVQKILIG 120
Db 61 IWDTAGOERYQITTKQYRRAGQIFLVYDISSERSYOHIMKMWSDVDEYAPGVQKILIG 120
QY 121 NKADEBQKRVGREQOQ-----QLAKYGMDFYETSACTNLNLIKESFTRLTVELVLAHR 174
Db 121 NKADEBQKRVGREQOQKPCPSQLAKYGMDFYETSACTNLNLIKESFTRLTVELVLAHR 180
QY 175 KELEGLMRASNELALAELEEEGKPEGPNSSKTCWC 212
Db 181 KELEGLMRASNELALAELEEEGKPEGPNSSKTCWC 218
RESULT 3
AAU17555
ID AAU17555 standard; Protein; 188 AA.
XX
XX AAU17555;

XX 07-NOV-2001 (first entry)
DT Novel signal transduction pathway protein, Seq ID 1120.
DE
XX
XX Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW Immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS;
XX acquired immune deficiency syndrome.
OS Homo sapiens.
XX
XX WO200154733-A1.
PN
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01312.
PF
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0231415.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-463460/50.
DR N-PSDB; AAS27472.
XX
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX
XX
XX Claim 1; SEQ ID NO 1120; 880pp; English.
XX
XX The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease and Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX

Query Match 75.3%; Score 832; DB 22; Length 188;
Best Local Similarity 97.6%; Pred. No. 6.2e-80;
Matches 160; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSSHISTIGVDVKMTIEVDGKVRQ 60
DB 24 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSSHISTIGVDVKMTIEVDGKVRQ 83
QY 61 IWDTAGQERYQITTKQYRRAGQIFLVYDISSERSYQHIMKWSDVDEYAPGVQKILIG 120
DB 84 IWDTAGQERYQITTKQYRRAGQIFLVYDISSERSYQHIMKWSDVDEYAPGVQXILIG 143
QY 121 NKADEEQKRVQREGQOLAKKEYGDMFYETSACTNLNLIKESFTR 164
DB 144 NKADEEQKRVQREGQOLAKKEYGDMFYETSACTNLNLIKESFTR 187

RESULT 4
ABB70670
ID ABB70670 standard; Protein; 204 AA.
XX

AC ABB70670;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 38802.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL14773.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 38802; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 204 AA;
SQ
Query Match 49.5%; Score 546.5; DB 22; Length 204;
Best Local Similarity 49.8%; Pred. No. 1.4e-49;
Matches 105; Conservative 47; Mismatches 50; Indels 9; Gaps 3;

QY 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSSHISTIGVDVKMTIEVDGKVRQ 59
DB 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSSHISTIGVDVKMTIEVDGKVRQ 60
QY 60 QIWDTAGQERYQITTKQYRRAGQIFLVYDISSERSYQHIMKWSDVDEYAPGVQKILI 119
DB 61 QIWDTAGQERYQITTKQYRRAGQIFLVYDISSERSYQHIMKWSDVDEYAPGVQKILI 120
QY 120 GNKADEEQKRVQREGQOLAKKEYGDMFYETSACTNLNLIKESFTRTELTLVLQAHKELEG 179
DB 121 GNKCDMTKRVVKNKGEAIAHEGIRFEMETSAKSNINIERAFCELAELD----KTSG 176
QY 180 LEMRASNELALAELEECKPEGPANSSKTC 210
DB 177 RESAENQERVIIDRRNQEKAP----GYSKCC 203

RESULT 5
ABB11916
ID ABB11916 standard; peptide; 213 AA.
XX
XX ABB11916;
AC

XX 11-JAN-2002 (first entry)
XX Human rab8 homologue, SEQ ID NO:2286.
XX
XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnary; antiulcer.
XX
XX Homo sapiens.
XX WO200157188-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US03800.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX WPI: 2001-457740/49.
XX N-PSDB; ABA09160.
XX
XX Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
XX Claim 20; Page 276; 1963pp; English.
XX
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth, factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to

CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention.
XX
XX Sequence 213 AA;
SQ
Query Match 49.3%; Score 545; DB 22; Length 213;
Best Local Similarity 52.9%; Pred. No. 2.1e-49;
Matches 99; Conservative 45; Mismatches 43; Indels 0; Gaps 0;
QY 1 MAKQYDLVFLRLLLIGDSGVGKTCCLCRFTDNFHHSHISTIGVDFKMKTIYDGVKIRIQ 60
Db 7 MAKTYDLFLKLLIGDSGVGKTCVLFPSFEDAFNSFTTIGIDFKIRTIELDKRKIKLQ 66
QY 61 IWDTAGQERYQITTKQYRRAGQIFLYVDISSERSQHIWKVSDVDEYAPGVQKILIG 120
Db 67 IWDTAGQERFRTTAYYRGAMGLVYDITNEKSFEDNIRNINTEEHASADVEMKILG 126
QY 121 NKADEBOKRQVREGOQOLAKEYGMDFYETSACTNLIKESFTRLTFLVLAQHRKELEGL 180
Db 127 NKCDVNDKRVSKERGEKALDYGKFMETSAKANINVENAFETLARDIKAKMDKKLEGN 186
QY 181 RMRASNE 187
Db 187 SPOGSNQ 193
RESULT 6
ID ABP41333 standard; Protein; 221 AA.
XX
AC ABP41333;
XX
DT 23-AUG-2002 (first entry)
XX
DE Human ovarian antigen HCGMA67, SEQ ID NO:2465.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive.
XX
OS Homo sapiens.
XX
XX WO200200677-A1.
XX
XX 03-JAN-2002.
XX
XX 07-JUN-2001; 2001WO-US18569.
XX
XX 07-JUN-2000; 2000US-209467P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Blirse CE, Rosen CA;
XX WPI: 2002-147878/19.
XX N-PSDB; ABQ54410.
XX
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
XX useful in the prevention, treatment and diagnosis of cancer (e.g.
XX ovarian cancer), immune disorders, cardiovascular disorders and
XX neurological diseases -

XX PS Claim 11; SEQ ID NO 2465; 2922pp; English.

XX CC The invention relates to 2175 novel human ovarian antigens (ABP41054 -

CC ABP4328) and to cDNAs encoding them (ABQ54131-ABQ56305), and also

CC encompasses polypeptides 90% identical and polynucleotides 95% identical

CC to the sequences of the invention. The invention additionally relates to

CC recombinant vectors and host cells comprising human ovarian antigen

CC polynucleotides, antibodies against human ovarian antigens, and the use

CC of ovarian antigen polynucleotides and polypeptides in diagnosing,

CC treating, prognosing or preventing various ovary and/or breast-related

CC disorders. Such conditions include ovarian cancer and breast cancer, and

CC metastatic tumours of ovarian or breast origin, reproductive system

CC disorders (e.g., infertility, disorders of pregnancy, anovulation,

CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine

CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic

CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and

CC vaginitis), immune disorders (e.g., congenital and acquired

CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),

CC blood-related disorders (e.g., anaemia), cardiovascular disorders,

CC respiratory disorders, neurological disorders, gastrointestinal disorders

CC and urinary system disorders. Ovarian antigen polypeptides and

CC polynucleotides may also be used in screening for compounds which

CC modulate ovarian antigen expression or activity. The polynucleotides may

CC further be used for gene therapy, chromosome mapping, in the

CC identification of individuals and in forensic analysis, and the

CC polypeptides may be used as food additives or to prepare antibodies

CC useful in disease diagnosis, drug targeting and phenotyping. The present

CC sequence represents a human ovarian antigen of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 221 AA;

Query Match 48.9%; Score 540; DB 23; Length 221;

Best Local Similarity 52.7%; Pred. No. 7.5e-49;

Matches 98; Conservative 45; Mismatches 43; Indels 0; Gaps 0;

Qy 2 AKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSHSHISTIGVDFKMTIEVDGKIVRIQI 61

Db 16 AKTYDLFRLLIGDSGVGKTCCLCRFTDNEFHSHSHISTIGVDFKMTIEVDGKIVRIQI 75

Qy 62 WDTAGQERYQTITKQYRRAGQIFLVYDTSERSYQHIMKMWSDVDEYAPGVQKILIGN 121

Db 76 WDTAGQERYQTITKQYRRAGQIFLVYDTSERSYQHIMKMWSDVDEYAPGVQKILIGN 135

Qy 122 KADEEQKRVQREGQQLAKKEYGMDFYETSACTNLNKESTRLTELVLQAHKKEGLR 181

Db 136 KCDVNDKRVSKERGEKALDYGIKFMETSAKANINVENAFPTLARDIKAKMDKKLEGS 195

Qy 182 MRASNE 187

Db 196 PQGSNQ 201

RESULT 7

AAB58196

ID AAB58196 standard; Protein; 246 AA.

XX AC AAB58196;

XX DT 14-MAR-2001 (first entry)

XX DE Lung cancer associated polypeptide sequence SEQ ID 534.

XX KW Human; lung cancer associated protein; neuroprotective; cytostatic;

XX KW cardioactive; immunomodulatory; muscular active; vulnerary;

XX KW gastrointestinal; nephrotropic; antiinfective; gynecological;

XX KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;

XX KW proliferative disorder; wound healing; infectious disease.

XX OS Homo sapiens.

XX PN WO200055180-A2.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US05918.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

FA (ROSE/) ROSEN C A.

Ruben SM;

WPI: 2000-587514/55.

N-PSDB; AAF18072.

Lung cancer associated gene sequences, referred to as lung cancer

PT antigens, useful for treatment, prevention, and diagnosis of disorders

PT such as lung cancer -

XX Claim 11; Page 1027-1028; 1425pp; English.

CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer

CC associated proteins represented in AAB58106 - AAB58548. Lung cancer

CC associated proteins and polynucleotide sequences, their agonists, and

CC antagonists may have neuroprotective; cytostatic; cardioactive;

CC immunomodulatory; muscular active general; vulnerary; gastrointestinal

CC general; nephrotropic; antiinfective; gynecological; or antibacterial

CC activity. The invention also includes antibodies specific for the

CC protein or polynucleotide sequences. The lung cancer associated

CC polynucleotide sequences may be used for detection of lung cancer,

CC chromosome identification, as chromosome markers, and for numerous other

CC diagnostic or research purposes. The proteins may be used to treat

CC disorders such as neural, immune, muscular, reproductive,

CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative

CC disorders. The proteins may also be used in the treatment of wounds and

CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and

CC peptide AAB58549 are used in the course of the invention for the

CC identification and characterisation of the polynucleotide and protein

CC sequences.

XX SQ Sequence 246 AA;

Query Match 48.1%; Score 532; DB 21; Length 246;

Best Local Similarity 47.6%; Pred. No. 6.2e-48;

Matches 100; Conservative 49; Mismatches 43; Indels 18; Gaps 3;

Qy 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSHSHISTIGVDFKMTIEVDGKIVRIQ 60

Db 44 MAKAYDHLFKLLIGDSGVGKTCCLIRFAEDNFNNTYISTIGIDFKIRTVDIEGKKIKLQ 103

Qy 61 IWDTAGQERYQTITKQYRRAGQIFLVYDTSERSYQHIMKMWSDVDEYAPGVQKILIG 120

Db 104 VMDTAGQERFKTITATTYYRGAMGIIIVDYDITDEKSFENIQNMWMSIKENASAGVERLLIG 163

Qy 121 NKADEEQKRVQREGQQLAKKEYGMDFYETSACTNLNKESTRLTELVLQAHKKEGLR 180

Db 164 NKCDMEARKVKQEQADKLAREHGIRFETSAKSMNVNDEAFSLARDIL-----LKSG 217

Qy 181 RMRASNELALAELEEEGKPEGPANSSKTC 210

Db 218 GRRSGN-----GNKP--PSTD LKTC 235

RESULT 8

AAB19165

ID AAB19165 standard; Protein; 200 AA.

XX AC AAB19165;

XX DT 19-FEB-2001 (first entry)

XX OS

```
DE Amino acid sequence of human RAB10.
KW RAB protein; GTPase; GTP binding; gene therapy; cancer.
XX
OS Homo sapiens.
XX
PN WO200058464-A2.
XX
PD 05-OCT-2000.
XX
PF 13-MAR-2000; 2000WO-US06330.
XX
PR 25-MAR-1999; 99US-0126083.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Allen M, Abel K, McIntosh B, Vega R, Rutter M, Buckler AJ;
XX
DR WPI; 2000-547233/62.
XX
DR N-PSDB; AAA96887.
XX
XX
XX Novel isolated nucleic acid encoding a mammalian RAB protein useful for
PT identifying homologous or related genes, in producing composition that
PT modulates expression or function of RAB for cancer therapy -
XX
XX Claim 2; Page 39; 58pp; English.
XX
XX The present sequence represents a mammalian RAB protein. RAB proteins
CC constitute the largest family of small GTPases, with over 40 identified
CC isoforms. RAB proteins contain four highly conserved peptide sequences
CC involved in GTP binding and hydrolysis. Compositions comprising RAB
CC nucleic acid are useful for identifying homologous or related genes,
CC in producing compositions that modulate the expression or function of
CC RAB, for gene therapy, mapping functional regions of the protein and
CC in studying associated physiological pathways. In addition, modulation
CC of the gene activity in vivo is used for prophylactic and therapeutic
CC purposes, such as treatment of cancer, and identification of cell type
CC based on expression. The DNA may also be used to identify expression of
CC the gene in a biological specimen.
XX
XX Sequence 200 AA;
XX
XX Query Match 48.0%; Score 530.5; DB 21; Length 200;
XX Best Local Similarity 55.8%; Pred. No. 6.6e-48;
XX Matches 96; Conservative 46; Mismatches 29; Indels 1; Gaps 1;
XX
QY 1 MAKQ-YDVLFRLLIGDSGVGKTCVLRFDNEFHSHSHSTIGVDFKMTIEVDGIKVI 59
Db 1 MAKTYDLLFKLLIGDSGVGKTCVLRFDNEFHSHSHSTIGVDFKMTIEVDGIKVI 60
QY 60 QIWDTAGQERYQITTKQYRRAGIFLVYDIDISSERSYQHIMKWSVDVDEYAPGVQKILI 119
Db 61 QIWDTAGQERFHITTSYRGAMGIMLVYDITNGKSFENISKWLRNIDEHANEDVERMLL 120
QY 120 GNKADEQKRVGREGQQQLAKEYGMDFYETSACTNINIKESFTRLTELVLQ 171
Db 121 GNKCDMDKRVVPKGEQIAREHGIRFFETSAKANINIEKAFLLTAEIDL 172
XX
XX RESULT 9
XX AAB09979
XX ID AAB09979 standard; Protein; 200 AA.
XX
XX AAB09979;
XX
XX 19-OCT-2000 (first entry)
XX
XX Human Rab10 protein.
XX
XX Human; Rab10; Yap/Rab family.
XX
XX Homo sapiens.
XX
XX
```

```
PN CN1249345-A.
XX
PD 05-APR-2000.
XX
PF 28-SEP-1998; 98CN-0121911.
XX
PR 28-SEP-1998; 98CN-0121911.
XX
PA (UYFU-) UNIV FUDAN.
XX
PI Yu L, Tu Q, Gao J;
XX
DR WPI; 2000-400723/35.
XX
DR N-PSDB; AAA40104.
XX
XX Preparation of human gene coding sequence, its encoded polypeptide -
XX Claim 2; Page 18; 23pp; Chinese.
XX
XX This invention describes a novel human Rab10 cDNA sequence. The protein
CC coded by said sequence belongs to Yap/Rab family and is the homolog of
CC mouse Rab10. The present invention also relates to the polypeptide coded
CC by said nucleotide sequence and the application and preparing process
CC of said polynucleotide and said polypeptide. This sequence represents the
CC human Rab10 protein described in the method of the invention.
XX
XX Sequence 200 AA;
XX
XX Query Match 48.0%; Score 530.5; DB 21; Length 200;
XX Best Local Similarity 55.8%; Pred. No. 6.6e-48;
XX Matches 96; Conservative 46; Mismatches 29; Indels 1; Gaps 1;
XX
QY 1 MAKQ-YDVLFRLLIGDSGVGKTCVLRFDNEFHSHSHSTIGVDFKMTIEVDGIKVI 59
Db 1 MAKTYDLLFKLLIGDSGVGKTCVLRFDNEFHSHSHSTIGVDFKMTIEVDGIKVI 60
QY 60 QIWDTAGQERYQITTKQYRRAGIFLVYDIDISSERSYQHIMKWSVDVDEYAPGVQKILI 119
Db 61 QIWDTAGQERFHITTSYRGAMGIMLVYDITNGKSFENISKWLRNIDEHANEDVERMLL 120
QY 120 GNKADEQKRVGREGQQQLAKEYGMDFYETSACTNINIKESFTRLTELVLQ 171
Db 121 GNKCDMDKRVVPKGEQIAREHGIRFFETSAKANINIEKAFLLTAEIDL 172
XX
XX RESULT 10
XX AAB95340
XX ID AAB95340 standard; Protein; 200 AA.
XX
XX AAB95340;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:17619.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
```


PD 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR N-PSDB; AAS71453.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 37625; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 218 AA;
Query Match 47.9%; Score 529.5; DB 22; Length 218;
Best Local Similarity 55.8%; Pred. No. 9.6e-48;
Matches 96; Conservative 46; Mismatches 29; Indels 1; Gaps 1;
Qy 1 MAKO-YDVLFRLLIGDSGVGKTCCLCRFTDNEPHSSHISTIGYDFRMKTIKIEVDGKVR 59
Db 19 MAKRTYDLFLKLLIGDSGVGKTCVLFPSDDAFNTFISTIGDFKIKTELQGGKKIL 78
Qy 60 QIWDAGQERYQTITTKQYIRRAQGIIFLVYDTSERSYQHIMKWYSDVDEVAPEGVQKILI 119
Db 79 QIWDAGQERFHTTTSYRGAMGIMLVYDITNGKSPENISKVLINIDEHANEDVERMLL 138
Qy 120 GNKADEQKRVGREGQQQLAKEYGMDFFYETSACTNLNIKESFTRLTFLVLAHRELEGL 171
Db 139 GNKCDMDKRVVPKGGQIAREHGIRFFETSAKANIINIEKAFLTAEIDILR 190
RESULT 13
ABB71647
ID ABB71647 standard; Protein; 207 AA.
XX ABB71647;
XX ABB71647;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 41733.
XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL15750.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure; SEQ ID NO 41733; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB27072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 207 AA;
Query Match 47.8%; Score 528.5; DB 22; Length 207;
Best Local Similarity 53.8%; Pred. No. 1.1e-47;
Matches 100; Conservative 41; Mismatches 40; Indels 5; Gaps 1;
Qy 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEPHSSHISTIGYDFRMKTIKIEVDGKVR 60
Db 1 MAKTYDYLFRLLIGDSGVGKTCILFRSDDAFNTFISTIGDFKIRTIENDKKIKLQ 60
Qy 61 IWDTAGQERYQTITTKQYIRRAQGIIFLVYDTSERSYQHIMKWYSDVDEVAPEGVQKILIG 120
Db 61 IWDTAGQERFHTTTSYRGAMGIMLVYDITNGKSPENIKWIRNIEENASADVERKMLL 120
Qy 121 NKADEQKRVGREGQQQLAKEYGMDFFYETSACTNLNIKESFTRLTFLVLAHRELEGL 180
Db 121 NKGLTDRKQVSRGEQQLAIEYGIKFMETSAKASINVEEAF-----LTASDIKATEK 175
Qy 181 RMRASN 186
Db 176 RMEANN 181
RESULT 14
AAG67154
ID AAG67154 standard; Protein; 207 AA.
XX AAG67154;
XX AAG67154;
XX 13-NOV-2001 (first entry)
XX Amino acid sequence of human 27423 G-protein.
XX Human; G-protein; 32705; 23224; 27423; 32700; 32712; lung disorder;
XX congenital anomaly; pulmonary congestion; oedema; haemorrhage;
KW

KW adult respiratory distress syndrome; Goodpasture's syndrome;
 KW chronic obstructive pulmonary disease; asthma; pulmonary hypertension;
 KW liver disorder; hepatic injury; jaundice; cholestasis; viral hepatitis;
 KW cirrhosis; Wilson's disease; autoimmune hepatitis; hepatic failure;
 KW brain disorder; hypoxia; cerebral ischemia; intracranial haemorrhage;
 KW acute meningitis; Parkinson's disease; Alzheimer's disease; glioma;
 KW chronic bacterial meningoencephalitis; multiple sclerosis;
 KW amyotrophic lateral sclerosis; stroke; Huntington's disease.
 XX Homo sapiens.
 OS
 XX
 PN WO200164887-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 27-FEB-2001; 2001WO-US06292.
 XX
 PF 29-FEB-2000; 2000US-0185606.
 XX
 PR (MILL-) MILLENNIUM PHARM INC.
 XX
 PA Meyers RA;
 PI
 XX WPI; 2001-550182/61.
 DR N-PSDB; AAG67154.
 XX
 XX Novel human small G-protein polypeptides and polynucleotides for
 PT treating lung disorders, liver disorders and brain disorders -
 PT
 XX
 PS Claim 8; Fig 15; 151pp; English.
 XX
 CC The present sequence represents a human G-protein. The specification
 CC describes 32705, 23224, 27423, 32700 or 32712 small G-proteins. The
 CC G-protein polypeptides and polynucleotides are useful as a target for
 CC diagnosis and treatment of G-protein mediated or related disorders,
 CC and for identifying agonists and antagonists for diagnosis and
 CC treatment. They are useful for treating disorders of lung (e.g.
 CC congenital anomalies, pulmonary congestion, oedema, adult respiratory
 CC distress syndrome, haemorrhage, chronic obstructive pulmonary disease,
 CC asthma, Goodpasture's syndrome and pulmonary hypertension), liver
 CC (e.g. hepatic injury, jaundice, cholestasis, viral hepatitis, cirrhosis,
 CC Wilson's disease, autoimmune hepatitis and hepatic failure), and
 CC brain (e.g. hypoxia, cerebral ischemia, intracranial haemorrhage, acute
 CC meningitis, Parkinson's disease, Alzheimer's disease, gliomas, chronic
 CC bacterial meningoencephalitis, multiple sclerosis, amyotrophic lateral
 CC sclerosis, stroke and Huntington's disease).
 XX
 SQ Sequence 207 AA;
 Query Match 47.8%; Score 528.5; DB 22; Length 207;
 Best Local Similarity 54.0%; Pred. No. 1.1e-47;
 Matches 95; Conservative 47; Mismatches 33; Indels 1; Gaps 1;
 QY 1 MAKQYDLFRLLIGDSGVGKTCLLCRFTDNEFHSHSTIGVDFKMKTEVDGKVRQ 60
 DB 1 MAKTYDLFKLLIGDSGVGKTCLLFRSEDAFTNTFISTIGIDFKIRTELDGKKIKLQ 60
 QY 61 IWDTAGQERYQTITKQYRRAGQIFLYVDISSERSYQHIMKWSDVDEYAPGVQKILIG 120
 DB 61 IWDTAGQERFTTITAYRGAMGIMLYDITNEKSPDNKNWIRNTEEHASSDVERMILG 120
 QY 121 NKADEBQKROVGRGOQQLAKKEYGMDFYETSACTNLIKES-FTRLTELVLQAHRK 175
 DB 121 NKCDMNDKROVSKERGEKLAIDYGIKFLFETSAKSSANVEAEFTLARDIMTKLNKR 176
 RESULT 15
 AAB92628
 ID AAB92628 standard; Protein; 207 AA.
 XX
 AC AAB92628;
 XX
 DT 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:10930.
 DE
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KW
 XX Homo sapiens.
 OS
 XX EP1074617-A2.
 PN
 XX 07-FEB-2001.
 PD
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 XX
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 XX WPI; 2001-318749/34.
 DR
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PT
 XX
 PS Claim 8; SEQ ID 10930; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 207 AA;
 Query Match 47.8%; Score 528.5; DB 22; Length 207;
 Best Local Similarity 54.0%; Pred. No. 1.1e-47;
 Matches 95; Conservative 47; Mismatches 33; Indels 1; Gaps 1;
 QY 1 MAKQYDLFRLLIGDSGVGKTCLLCRFTDNEFHSHSTIGVDFKMKTEVDGKVRQ 60
 DB 1 MAKTYDLFKLLIGDSGVGKTCLLFRSEDAFTNTFISTIGIDFKIRTELDGKKIKLQ 60
 QY 61 IWDTAGQERYQTITKQYRRAGQIFLYVDISSERSYQHIMKWSDVDEYAPGVQKILIG 120
 DB 61 IWDTAGQERFTTITAYRGAMGIMLYDITNEKSPDNKNWIRNTEEHASSDVERMILG 120
 QY 121 NKADEBQKROVGRGOQQLAKKEYGMDFYETSACTNLIKES-FTRLTELVLQAHRK 175
 DB 121 NKCDMNDKROVSKERGEKLAIDYGIKFLFETSAKSSANVEAEFTLARDIMTKLNKR 176

Search completed: November 17, 2002, 20:38:26
Job time : 68 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2002, 20:39:14 ; Search time 2665 Seconds
(without alignments)
2315.120 Million cell updates/sec

Title: US-09-817-198A-2

Perfect score: 1105

Sequence: 1 MAKQYDVLFRLLIGDSGVG.....LEEERKPEGPANSSKTCWC 212

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09817198/runat_13112002_135207_7033/app_query.fasta_1.391
-DB=genembl -QWMT=fascap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09817198.ecgn.1.1616.9runat_13112002_135207_7033 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vi:*
30: em.htg_hum:*
31: em.htg_inv:*
32: em.htg_other:*
33: em.htg_mus:*
34: em.htg_pln:*
35: em.htg_rod:*
36: em.htg_mam:*
37: em.htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB Length	ID	Description
1	1105	100.0	1054	6	AX399903	AX399903 Sequence
2	1090	98.6	3139	10	BC027769	BC027769 Mus muscu
3	1077	97.5	945	10	RATRA15X	M83679 Sprague-Daw
4	798	72.2	2560	10	BC013790	BC013790 Mus muscu
5	546.5	49.5	1613	3	AB006189	AB006189 Drosophil
6	546.5	49.5	2349	3	AY060425	AY060425 Drosophil
7	545	49.3	624	9	AF498943	AF498943 Homo sapi
8	545	49.3	660	9	HSMRAB8	X56741 H.sapiens m
9	545	49.3	760	4	CFRAB8	X56385 Canine rab8
10	545	49.3	2048	9	BC002977	BC002977 Homo sapi
11	545	49.3	2818	9	AK025165	AK025165 Homo sapi
12	541	49.0	1337	10	BC019990	BC019990 Mus muscu
13	539.5	48.8	765	5	DYGORA2	M38391 Discoppyge o
14	534.5	48.4	1211	9	BC000799	BC000799 Homo sapi
15	534.5	48.4	1238	9	HSRAB13	X75593 H.sapiens m
16	532	48.1	612	9	AF498948	AF498948 Homo sapi
17	531.5	48.1	840	5	DYGORAL	M38390 Discoppyge o
18	530.5	48.0	603	9	AF297660	AF297660 Homo sapi
19	530.5	48.0	603	9	AF498945	AF498945 Homo sapi
20	530.5	48.0	888	4	CFRAB10	X56387 Canine rab1
21	530.5	48.0	897	10	AF035646	AF035646 Mus muscu
22	530.5	48.0	1029	9	BC000896	BC000896 Homo sapi
23	530.5	48.0	3164	9	AF106681	AF106681 Homo sapi
24	530.5	48.0	3533	9	AK023223	AK023223 Homo sapi
25	529.5	47.9	954	8	ATHARA3	D01025 Arabidopsis
26	529.5	47.9	969	8	AY042795	AY042795 Arabidops
27	529.5	47.9	1143	8	AY035132	AY035132 Arabidops
28	528.5	47.8	624	6	AX236078	AX236078 Sequence
29	528.5	47.8	740	10	RNU53475	U53475 Rattus norv
30	528.5	47.8	759	10	AF525280	AF525280 Rattus no
31	528.5	47.8	1128	9	BC020654	BC020654 Homo sapi
32	528.5	47.8	1161	6	AX236076	AX236076 Sequence
33	528.5	47.8	1265	9	AB038995	AB038995 Homo sapi
34	528.5	47.8	2048	3	AY069671	AY069671 Drosophil
35	528.5	47.8	2210	3	D84347	D84347 Drosophila
36	528.5	47.8	2497	6	AX285074	AX285074 Sequence
37	528.5	47.8	2497	6	AX285080	AX285080 Sequence
38	528.5	47.8	2497	6	AX285089	AX285089 Sequence
39	528.5	47.8	2497	6	AX285091	AX285091 Sequence
40	528.5	47.8	3077	9	AK001111	AK001111 Homo sapi
41	528.5	47.8	3110	9	HSM804678	AL833365 Homo sapi
42	526.5	47.6	1113	8	LJRBABE	Z73948 L.japonicus
43	526.5	47.6	3521	9	HSM801620	AL36650 Homo sapi
44	526	47.6	638	10	S53270	S53270 MEL-RAS-rel
45	526	47.6	1118	10	BC027214	BC027214 Mus muscu

ALIGNMENTS

RESULT 1

```
AX399903      1054 bp      DNA      linear      PAT 06-JUN-2002
LOCUS          Sequence 74 from Patent WO0218424.
DEFINITION    AX399903
ACCESSION     AX399903
VERSION       AX399903.1 GI:21336157
KEYWORDS      human.
SOURCE        Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1
AUTHORS      Tang, Y.T., Asundi, V., Zhou, P., Xue, A.J., Ren, F., Zhang, J.,
              Wang, J.R., Zhao, Q.A., Wang, D., Liu, C., Drmanac, R.T. and Wehrman, T.
TITLE         Nucleic acids and polypeptides
              Patent: WO 0218424-A 74 07-MAR-2002;
              HYSEQ, INC. (US)
FEATURES      Location/Qualifiers
source        1..1054
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              81..719
              /note="unnamed protein product"
              /codon_start=1
              /protein_id="CAD33347.1"
              /db_xref="GI:21336158"
              /translation="MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSHSTIGVD
              FMKKTIEVDGIRVRIQIMDTAGQRYQITTKYRAQGIPLVYDISSERSYQHIMKW
              VSDVPEYAPGVQKILIGNKDEOKROVGEOOLAKYMGDFEYSACFNLIKE
              STRTELVLQHRLEGLMRASNELALAELEEKGPEGPANSSKTCWC"
BASE COUNT    240 a 314 g 304 g 195 t 1 others
ORIGIN
Alignment Scores:
Pred. No.:      2,15e-100      Length:      1054
Score:          1105.00      Matches:      212
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:              Gaps:      0

US-09-817-198a-2 (1-212) x AX399903 (1-1054)

Qy 1 MetalLysGlnTyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerGlyValGly 20
|||||
Db 81 ATGCGAAGCAGTACGATGCTGCTCCGCTGCTGCTGATCGGGGACTCCGGGTGGC 140
|||||
Qy 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisIleSerThr 40
|||||
Db 141 AAGACCTGCCTGCTGTCGCCGCTTACCAGCAACGAGTTCCTCCTCGCATCTCCACC 200
|||||
Qy 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGln 60
|||||
Db 201 ATCGGTGTTGACTTTAAGATGAACACCATAGAGGTAGACGCGATCAAGTGGGATACAG 260
|||||
Qy 61 IleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrArgArg 80
|||||
Db 261 ATCTGGGACACTGCGAGGCGGAGAGATACACCATCAACAGCAGTACTATCGCGCG 320
|||||
Qy 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
|||||
Db 321 GCCCAGGGGATATTTTGGCTATGACATATAGACGAGCGGCTCTTACCAGCATCATG 380
|||||
Qy 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
|||||
Db 381 AAGTGGGTGTCAGTGCCTGGATGAGTACGCACCAAGCGCTCCAGAAGATCCTTATTGGG 440
|||||
Qy 121 AsnLysAlaAspGluGluGlnLysArgGlnValGlyArgGluGlnGlnGlnLeuAla 140
|||||
Db 441 AATAAGCGCTGATGAGGACGAGAAACGGCAGCTGGGAGAGACGAGCGAGCTGGCG 500
|||||
Qy 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
|||||
Db 501 AAGGATGATGCGCATGGACTTCTTGAACAAAGTCCTGCACCAACCTCAACATTAAGAG 560
|||||
```

```
Qy 161 SerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGlyLeu 180
|||||
Db 561 TCATTACACGCGTCTGACAGAGCTGGTCTGCGAGGCCCATAGGAAGAGCTGGAAGCCCTC 620
|||||
Qy 181 ArgMetArgAlaSerAsnGluLeuAlaLeuAlaGluLeuGluGluGluGlyLysPro 200
|||||
Db 621 CGATGCGTGGCCAGCAATGAGTTGGCAGCTGGCAGAGCTGGAGGAGGAGGAGCAACCC 680
|||||
Qy 201 GluGlyProAlaAsnSerSerLysThrCysTrpCys 212
|||||
Db 681 GAGGGCCACGCAACTCTTCGAAACCTGCTGGTGC 716
|||||
RESULT 2
BC027769      3139 bp      mRNA      linear      ROD 07-AUG-2002
LOCUS          Mus musculus, clone MGC:38375 IMAGE:5345297, mRNA, complete cds.
DEFINITION    BC027769
ACCESSION     BC027769
VERSION       BC027769.1 GI:20380721
KEYWORDS      MGC.
SOURCE        house mouse.
ORGANISM      Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              1 (bases 1 to 3139)
              Strausberg,R.
              Direct Submission
              Submitted (08-APR-2002) National Institutes of Health, Mammalian
              Gene Collection (MGC), Cancer Genomics Office, National Cancer
              Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
              USA
              NIH-MGC Project URL: http://mgc.nci.nih.gov
              Contact: MGC help desk
              Email: cgabbs-remail.nih.gov
              Tissue Procurement: Jeffrey Green M.D.
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: National Institutes of Health Intramural
              Sequencing Center (NISC),
              Gaithersburg, Maryland;
              Web site: http://www.nisc.nih.gov/
              Contact: nisc.mgc@nih.gov
              Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
              Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
              Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
              Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
              Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
              McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
              Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
              Young, A., Zhang, L.-H. and Green, E.D.
              Clone distribution: MGC clone distribution information can be found
              through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
              Series: IRAK Plate: 55 Row: d Column: 5
              This clone was selected for full length sequencing because it
              passed the following selection criteria: Hexamer frequency ORF
              analysis.
FEATURES      Location/Qualifiers
source        1..3139
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /map="FVB/N"
              /clone="MGC:38375 IMAGE:5345297"
              /tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
              ductal carcinoma. 5 month old virgin mouse."
              /clone_lib="NCI_CGAP_Mam6"
              /lab_host="DH10B"
              /note="Vector: pCMV-SPORT6"
              208..846
              /codon_start=1
              /product="Unknown (protein for MGC:38375)"
              /protein_id="AAH27769.1"
              /db_xref="GI:20380722"
              CDS
```

/translation="MAKQYDVLFRLLIGDSGVGKTKCLLRCFTDNEFHSHSTIGVD
FKMKTIDVGGIKVRIQIWDTAGQERYQITIKQYRRAQGIIFLYDISSERSYQHIMKW
VSDVDEYAPGVQKILIGNKADEQKROVREGQOOLAKKEYGMDYFETSACTNLNIKE
SFTRLTELVLQAHKRLDGLRNASNELALAELEDECKPEGPANSSKTCWC"
BASE COUNT 723 a 847 c 878 g 691 t
ORIGIN

Alignment Scores:
Pred. No.: 2,5e-98 Length: 3139
Score: 1090.00 Matches: 208
Percent Similarity: 99.53% Conservative: 3
Best Local Similarity: 98.11% Mismatches: 1
Query Match: 98.64% Indels: 0
DB: 10 Gaps: 0

US-09-817-198a-2 (1-212) x BC027769 (1-3139)

QY 1 MetAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerglyValGly 20
|||||
Db 208 ATGGCGAAACAGTACGATGCTGTTCGGCTACTGCTGATCGGGACTCCGGGTTGGC 267
QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisIleSerThr 40
|||||
Db 268 AGACATGCTGCTGCGCGCTTCACGACACAGAGTTCCTCTCGCATATCTCCACC 327
QY 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGln 60
|||||
Db 328 ATCGGTGTGACTTTAAGATGAAGACTATCGATGACGGCATCAAACTGAGATACAG 387
QY 61 IleTyrAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
|||||
Db 388 ATTTGGGACACAGCAGCGGAGGAGGTACACAGACTATCACAAAGCAGTACTATCGGCA 447
QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrClnHisIleMet 100
|||||
Db 448 GCCCGAGGAATATTTTACTGACGACATGACGAGTACGCTCTCTATCAGCATATCATG 507
QY 101 LysTyrValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
|||||
Db 508 AAGTGGTCTGATGCTGATGATGCTCCAGAAAGAGTCCAGAAAGATCTTAATTTGGG 567
QY 121 AsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnGlnLeuAla 140
|||||
Db 568 AATAAGGCTGATGACAGAGCAGAAACGGCAGGTGGGAGAGAGCAGGGGAGCAGTGGCT 627
QY 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
|||||
Db 628 AAGGAGTACGGCATGGACTTCACGAAACAGTGCTGCACCAACCTTAATTAAGAAG 687
QY 161 SerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGlyLeu 180
|||||
Db 688 TCCTTCACCTGCTGACGGAGCTGGTGTGAGAGGCCACAGGAAGAGTGGATGGTCTC 747
QY 181 ArgMetArgAlaSerAsnGluLeuAlaLeuAlaGluLeuGluGluGlyLysPro 200
|||||
Db 748 CGAACACGTCGCCCAACGAGCTGCGACATGCGCGAGCTGGAGGAGCAGCAAGCAACCT 807
QY 201 GluGlyProAlaAsnSerSerLysThrCysTyrCys 212
|||||
Db 808 GAGGGCCCGCAAACTCTTCAAGAGACCTGCTGCTGTC 843

RESULT 3
RATRAB15X
LOCUS RATRAB15X 945 bp mRNA linear ROD 27-APR-1993
DEFINITION Sprague-Dawley (clone LRB9) RAB15 mRNA, complete cds.
ACCESSION M83679
VERSION M83679.1 GI:206536
KEYWORDS LMW GTP-binding protein.
SOURCE Rattus norvegicus (strain Sprague-Dawley) (library: LAMBDA ZAP11)
adult brain cDNA to mRNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine;

Rattus.
1 (bases 1 to 945)
Elferink, L.A., Anzal, K. and Scheller, R.H.
rab15, a novel low molecular weight GTP-binding protein
specifically expressed in rat brain
J. Biol. Chem. 267 (9), 5768-5775 (1992)
92210533
1313420
Location/Qualifiers
1. .945
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
/tissue_lib="LAMBDA ZAP11"
1. .945
/gene="RAB15"
220. .858
/gene="RAB15"
/codon_start=1
/product="RAB15"
/protein_id="AAA41995.1"
/db_xref="GI:206537"
/translation="MAKQYDVLFRLLIGDSGVGKTKCLLRCFTDNEFHSHSTIGVD
FKMKTIDVGGIKVRIQIWDTAGQERYQITIKQYRRAQGIIFLYDISSERSYQHIMKW
VSDVDEYAPGVQKILIGNKADEQKROVREGQOOLAKKEYGMDYFETSACTNLNIKE
SFTRLTELVLQAHKRLDGLRNASNELALAELEDECKTEGPANSSKTCWC"

BASE COUNT 222 a 259 c 292 g 172 t
ORIGIN
Alignment Scores:
Pred. No.: 1,12e-97 Length: 945
Score: 1077.00 Matches: 207
Percent Similarity: 98.58% Conservative: 2
Best Local Similarity: 97.64% Mismatches: 3
Query Match: 97.47% Indels: 0
DB: 10 Gaps: 0
US-09-817-198a-2 (1-212) x RATRAB15X (1-945)

QY 1 MetAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerglyValGly 20
|||||
Db 220 ATGGCGAAACAGTACGATGCTGTATTTTCGGCTACTGCTGATCGGGACTCCGGGTTGGC 279
QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisIleSerThr 40
|||||
Db 280 AAGACTGCTGCTATGCGCTTCACCGACACAGAGTTCCTCTCGCATATCTCCACC 339
QY 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGln 60
|||||
Db 340 ATCGGTGTGACTTTAAGATGAAACCATCGAAGTAGAGCGGCATCAAGTAGGATACAG 399
QY 61 IleTyrAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
|||||
Db 400 ATTTGGGACACAGCAGCGGAGGAGGTACACAGACTATCACAAACAGTACTACCGCGCA 459
QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
|||||
Db 460 GCCCAGGGAATATTTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 519
QY 101 LysTyrValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
|||||
Db 520 AAGTGGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 579
QY 121 AsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnGlnLeuAla 140
|||||
Db 580 AATAAGGCTGACGAGAACAGAAACGGCAGGTGGGAGAGAGCAAGGGCAGCAGCTGGCT 639
QY 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
|||||
Db 640 AAGGAGTATGGATGGACTTCTACGAAACAGTGCCTGCACCAACCTTAACATTAAGAAG 699

```
QY 161 SerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGlyLeu 180
Db 700 TCCTTCACTCGTCTGACGAGCTGGTCTGCAGGCTCACAGAAAGAGCTGGATGGTCTC 759
QY 181 ArgMetArgAlaSerAsnGluLeuAlaLeuAlaGluLeuGluGluGlyLysPro 200
Db 760 CGAATATGTCACCAATAGACTCGCACTGGCCGAGCTGGAGGAAGACGAAGCAAAACT 819
QY 201 GluGlyProAlaAnSerSerLysThrCysTrpCys 212
Db 820 GAAGGCCCCAGCAAACTCTTCAAGACCTGCTGGTGC 855

RESULT 4
BC013790 2560 bp mRNA linear ROD 07-AUG-2002
LOCUS Mus musculus, clone MGC:6897 IMAGE:2655151, mRNA, complete cds.
DEFINITION BC013790
ACCESSION BC013790.1 GI:15489393
VERSION MGC.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 2560)
JOURNAL Strausberg,R.
Direct Submission
Submitted (04-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NTH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 5 Row: i Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, GenomeScan gene prediction.
FEATURES
source
1..2560
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N-3"
/clone="MGC:6897 IMAGE:2655151"
/tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month
old mouse. Taken by biopsy."
/clone_lib="NCI CGAP_Mam2"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
154..660
/codon_start=1
/product="Unknown (protein for MGC:6897)"
/protein_id="AAH13790.1"
/db_xref="GI:15489394"
/translation="MAKQDVLFRLLLGDSGVGKTCLLCRFTDNEFHSHSTIGVD
FMKMTIDVGIKVIQIMTGADEQRYQITKQYRAAGIFLVYDISSERSYQIHKW
VSDVDEYPAEGVKIILGNKADEQKRVGREGQQLAKEYGMDFEYTSACFFSFFS
FSTPSVPL"
BASE COUNT 614 a 648 c 719 g 579 t
ORIGIN
```

```
Alignment Scores:
Pred. No.: 1-65e-69 Length: 2560
Score: 798.00 Matches: 152
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.35% Mismatches: 0
Query Match: 72.22% Indels: 0
DB: 10 Gaps: 0

US-09-817-198a-2 (1-212) x BC013790 (1-2560)

QY 1 MetAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerGlyValGly 20
Db 154 ATGCGAAACAGTACGATGTCTGTCGCGCTACTGCTGATCGGGGACTCCGGGTGGC 213
QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisIleSerThr 40
Db 214 AAGACATGCTGCTGTGCGCTTCCACGACACAGAGTTCCTCCTCGCATATCTCCACC 273
QY 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGln 60
Db 274 ATCGGTGTTGACTTAAAGATGAAGACTATCGATGTAGACGGCATCAAGTGAGAAATACAG 333
QY 61 IleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
Db 334 ATTTGGGACACAGCAGCGGAGAGGTACCGACTATCACAAAGCAGTACTATCGGCGA 393
QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
Db 394 GCCCAGGGAATATTTTAGTCTACGACATTAGCAGTGAGCGCTCCTATCAGCATATCATG 453
QY 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
Db 454 AACTGGGTGAGTACGCTGAGTACGCTCCAGAGGAGTCCAGAGATCCTTAATTGGG 513
QY 121 AsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnGlnLeuAla 140
Db 514 AATAAGGCTGATGAAGACAGCAAAACGCGAGTGGGGAGAGAGCGGCGCAGCAGCTGGCT 573
QY 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCys 153
Db 574 AAGGATACGGCATGACCTTCTACGAACAAGTGCCTGC 612

RESULT 5
AB006189 1613 bp mRNA linear INV 07-AUG-1997
LOCUS Drosophila melanogaster mRNA for Rab10, complete cds.
DEFINITION AB006189
ACCESSION AB006189
VERSION AB006189.1 GI:2317271
KEYWORDS Rab10.
SOURCE Drosophila melanogaster (strain:Oregon R) head cDNA to mRNA.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE Satch,A.K., Tokunaga,F. and Ozaki,K.
AUTHORS Rab proteins of Drosophila melanogaster: novel members of the
TITLE Rab-protein family
JOURNAL FEBS Lett. 404 (1), 65-69 (1997)
MEDLINE 97228579
REFERENCE 2 (bases 1 to 1613)
AUTHORS Ozaki,K.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-1997) Koichi Ozaki, Osaka University, Graduate
School of Science, Department of Biology, 1-1 Machikaneyama,
Toyonaka, Osaka 560, Japan (E-mail:ozaki@bio.sci.osaka-u.ac.jp,
Tel:+81-6-850-5439, Fax:+81-6-850-5439)
FEATURES
Location/Qualifiers
1..1613
/organism="Drosophila melanogaster"
/strain="Oregon R"
/db_xref="taxon:7227"
```

gene	/tissue_type="head"	
	1. .1613	
	/gene="Drab10"	
	257. .871	
CDS	/gene="Drab10"	
	/codon_start=1	
	/product="Rab10"	
	/protein_id="BAA21744.1"	
BASE COUNT	/db_xref="GI:231727"	
	/translation="MAKTYDLFLKLLIGDSGVGKTCILFRFSDAFTSTFTIGI	
	DFKITYELRGKKIKLIQWDTAGQERFHTITTSYRGAMGIMLYVDITNEKSFENIV	
	WRNIDEHANEDVERKMILGNKCDMTDKRVNKEGEATAREHGIPTMETSAKSNINIE	
ORIGIN	RAFCEALAEILDKTSGRESAENQERVIDRRNQRKAPGYSKCCA"	
	460 a	365 c 430 g 358 t
	Alignment Scores:	
	Pred. No.:	7.7e-45 Length: 1613
Score:	546.50 Matches: 105	
	Percent Similarity: 72.04% Conservativeness: 47	
	Best Local Similarity: 49.76% Mismatches: 50	
	Query Match: 49.46% Indels: 9	
DB:	Gaps: 3	
	US-09-817-198a-2 (1-212) x AB006189 (1-1613)	
QY	1	MetAlaLysGln---TyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerGlyVal 19
Db	257	ATGCCAAGAAACCTACCATTTGCTCTTAACTGTTCTGCTGATCGGTATTCAGGAGTG 316
	20	GlyLysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisIleSer 39
	317	GGCAAGACGTGCATATTGTCGGTCTTCGGATGATGCATTCACGTCACCGTTCATATCG 376
QY	40	ThrIleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIle 59
	377	ACCATAGGCATCATTTCAAAATCAAAACAGTCGAGCTCGCGCGAAGATCAACGCTG 436
QY	60	GlnIleTyrAspThrAlaGlyGlnGluArgTyrGlnThrIlePheLysGlnTyrTyrArg 79
	437	CAATATGGACACCCCGCCAGGAGCGGTTCACACATACACCTCGTACTATCGA 496
QY	80	ArgAlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIle 99
	497	GGCGCCATGGCATAATGCTGCTATGACATAACGAACGAGAAGAGTTTCGAGAACATA 556
QY	100	MetLysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIle 119
	557	GTCAATGGTTACGGAATATAGACGACGACGCCACGAGGATGTGGAGAAGATGATCCTC 616
QY	120	GlyAspLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlyGlnGlnLeu 139
	617	GGCAACAAGTCGATATGACGGNACAGAGGGTGTCAACGAAGAGCGCGGCAAGCGATT 676
QY	140	AlaLysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLys 159
	677	GCCTGGTAACATGGCATTCGGTTATGTAAGAACATCCGCCAAGTCGAACATAAATCATCGAG 736
QY	160	GluSerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGly 179
	737	CGGCGCTTTCGAGCTGCGGAGGCAATCTGGAC-----AAGACATCAGGA 784
QY	180	LeuArgMetArgAlaSerAsnGluLeuAlaLeuAlaGluLeuGluGluGlyLys 199
	785	CGCGACTGCGCGAGAAATCAGGACCGCGTATTCATCGTCGCGGAACACGAGAGGCG 844
QY	200	ProGluGlyProAlaAsnSerSerLysThrCys 210
	845	CCG-----GGCTACACCAAGTGTGTC 865
LOCUS	RESULT 6	
	AY060425	
	2349 bp mRNA linear INV 05-NOV-2001	
	BASE COUNT 700 a 505 c 556 g 588 t	
ORIGIN	Alignment Scores:	
	Pred. No.:	1.22e-44 Length: 2349
	Score: 546.50 Matches: 105	
	Percent Similarity: 72.04% Conservativeness: 47	
DB:	Best Local Similarity: 49.76% Mismatches: 50	
	Query Match: 49.46% Indels: 9	
	Gaps: 3	
	US-09-817-198a-2 (1-212) x AY060425 (1-2349)	
DEFINITION	Drosophila melanogaster LD99986 full length cDNA.	
	ACCESSION AY060425	
	VERSION 1	
	GI:16648397	
KEYWORDS	FLI_CDNA	
	Drosophila melanogaster	
	Drosophila melanogaster	
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 2349)	
	Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J., Champer,M., Chavez,C., Dorsett,V., Farfan,D., Friese,E., George,R., Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M. and Celnikier,S.	
	Direct Submission	
	Submitted (29-OCT-2001) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA	
COMMENT	Sequence submitted by:	
	Berkeley Drosophila Genome Project	
	Lawrence Berkeley National Laboratory	
	Berkeley, CA 94720	
TITLE	This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (http://fruitfly.berkeley.edu) or send email to cdna@fruitfly.berkeley.edu.	
	Location/Qualifiers	
	1. .2349	
	/organism="Drosophila melanogaster"	
FEATURES	/strain="y; cn bw sp"	
	/db_xref="taxon:7227"	
	/map="19C1-19C1"	
	1. .2349	
gene	/gene="Rab10"	
	/note="alignment with genomic scaffold AE003572"	
	/db_xref="FLYBASE:FBgn0015789"	
	255. .869	
CDS	/gene="Rab10"	
	/note="Longest ORF"	
	/codon_start=1	
	/product="LD39986p"	
	/protein_id="AAL25464.1"	
	/db_xref="GI:16648398"	
	/db_xref="FLYBASE:FBgn0015789"	
	/translation="MAKTYDLFLKLLIGDSGVGKTCILFRFSDAFTSTFTIGI	
	DFKITYELRGKKIKLIQWDTAGQERFHTITTSYRGAMGIMLYVDITNEKSFENIV	
	WRNIDEHANEDVERKMILGNKCDMTDKRVNKEGEATAREHGIPTMETSAKSNINIE	
	RAFCEALAEILDKTSGRESAENQERVIDRRNQRKAPGYSKCCA"	
	700 a	505 c 556 g 588 t

Qy 1 MetAlaLysGln---TyrAspValLeuPheArgLeuLeuLeuLeuLeuGlyAspSerGlyVal 19
Db 255 ATGCCAAGAAACACCTACGATTTCTCTTAAACCTGTTCGTGATCGGTGATTCAGGAGTG 314
Qy 20 GlyLysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisIleSer 39
Db 315 GGCAGACGTCATATTGTCGGGTTCTCGGATGATCATCTCCAGTTCATATCG 374
Qy 40 ThrIleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIle 59
Db 375 ACCATAGGCATCGATTTCAAATCAAAACAGTCGAGTCGCGCGCAAGAAGATCAAGCTG 434
Qy 60 GlnIleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArg 79
Db 435 CAATATGGGACACCGCGCGGACGAGCGGTTCACACAGATTAACACCTCGTACTATCGA 494
Qy 80 ArgAlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIle 99
Db 495 GGCACCATGGCATATGCTGCTATGACATACAGACGAGAAGAGTTTCGAGAACATA 554
Qy 100 MetLysTrpValSerAspValAspLysAlaProGluGlyValGlnLysIleLeuIle 119
Db 555 GTCAATATGGTTACGGAATATAGACGACGCGCAACGAGGATGTGGAGAAGATCATCCTC 614
Qy 120 GlyAsnLysAlaAspGluGluGlnLysArgGlnValGlyArgGluGlnGlnLeu 139
Db 615 GGCACCAAGTGCATATGACGACGACAAAGGGTGTCAACAGGAGCGCGGCAAGCGAT 674
Qy 140 AlaLysGluTyrGlyMetAspPheThrGluThrSerAlaCysThrAsnLeuAsnIleLys 159
Db 675 GCCCGTGAACATGCATTCGGTTTATGGAACATCCGCCAAGTCCGAACATAAATCAG 734
Qy 160 GluSerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGly 179
Db 735 CGGCGCTTCGCGAGCTGGCGAGGCGCATCTCGAC-----AAGACATCAGGA 782
Qy 180 LeuArgMetArgAlaSerAsnGluLeuAlaLeuAlaGluLeuGluGluGluGlyLys 199
Db 783 CGCGAGTCGGCGGAATCAGGAGCGGTGATTATCATCGCCGGAACAGGAGGCGG 842
Qy 200 ProGluGlyProAlaAsnSerSerLysThrCys 210
Db 843 CCG-----GGCTACAGCAAGTGTGTCG 863
RESULT 7
AF498943
LOCUS
DEFINITION Homo sapiens small GTP binding protein RAB8 (RAB8) mRNA, complete
AF498943
VERSION AF498943.1 GI:20379061
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 624)
Publ.H.L. III, Ikeda,S.R. and Aronstam,R.S.
TITLE Homo sapiens RAB family small GTP binding protein RAB8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 624)
Publ.H.L. III, Ikeda,S.R. and Aronstam,R.S.
TITLE Direct Submission
JOURNAL Submitted (05-APR-2002) cDNA Resource Center, Guthrie Research
Institute, One Guthrie Square, Sayre, PA 18840, USA
FEATURES
source
1..624
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="brain"
1..624
/gene="RAB8"
1..624

BASE COUNT 188 a 147 c 165 g 124 t
ORIGIN
Alignment Scores:
Pred. No.: 3,4e-45 Length: 624
Score: 545.00 Matches: 99
Percent Similarity: 77.01% Conservative: 45
Best Local Similarity: 52.94% Mismatches: 43
Query Match: 49.32% Indels: 0
DB: 9 Gaps: 0
US-09-817-198A-2 (1-212) x AF498943 (1-624)
Qy 1 MetAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuLeuGlyAspSerGlyValGly 20
Db 1 ATGCCGAAGACCTACGATTACCTGTTCAAGCTGCTGCTGATCGGGGACTCGGGGGTGGG 60
Qy 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisIleSerThr 40
Db 61 AAGACCTGTGTCCTGTCGCGCTTCGCGAGGACGCCCTCACTCCACTTTATCTCCACC 120
Qy 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGln 60
Db 121 ATAGGAATTCACCTTTAAATATTAGACCATAGAGCTCGATGGCAAGAGATTTAAATCAG 180
Qy 61 IleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArg 80
Db 181 ATATGGGACACAGCGGTTCAGGAACGTTTCGGACGATCAACACGCGCTACTACAGGGGT 240
Qy 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
Db 241 GCAATGGGCATCATGCTGCTGATCATCATCAACACGAGAGAGTCTTCGACACATCCGG 300
Qy 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
Db 301 AACTGGATTTCGCAACATTTAGGAGCAGCGCTCTCGACAGGTCGAAAGATGATATCTCGG 360
Qy 121 AsnLysAlaAspGluGluGlnLysArgGlnValGlyArgGluGlnGlnGlnLeuAla 140
Db 361 AACAAAGTGTGATGTGAATGACAAAGAGACAAGTTTCCAGGAACGCGGAGAAAGCTGCC 420
Qy 141 LysGluTyrGlyMetAspPheThrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
Db 421 CTCGACATGGAATCAAGATTCATGAGACACCGCGGAGGACCAACATCAATGTGGAAT 480
Qy 161 SerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGlyLeu 180
Db 481 GCATTTTCTACTCTCGCCAGAGATCAAAAGCAAAATGGACAAAAATTTGGAAGGCAAC 540
Qy 181 ArgMetArgAlaSerAsnGlu 187
Db 541 AGCCCCCAGGGGAGCAACCCAG 561
RESULT 8
HSMRAB8
LOCUS
DEFINITION H.sapiens mRNA for rab8 gene.
ACCESSION X56741
VERSION X56741.1 GI:452317
KEYWORDS rab-related GTP-binding protein.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE AUTHORS TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 660) Zahraoui, A. Direct Submission
JOURNAL	Submitted (26-NOV-1990) A. Zahraoui, INSERM-U 248, 10 AVENUE DE VERDUN, 750-10 PARIS, FRANCE
REFERENCE AUTHORS TITLE	2 (bases 1 to 660) Zahraoui, A., Joberty, G., Arpin, M., Fontaine, J.J., Hellio, R., Tavtavian, A. and Louvard, D. A small rab GTPase is distributed in cytoplasmic vesicles in non polarized cells but colocalizes with the tight junction marker ZO-1 in polarized epithelial cells
JOURNAL	J. Cell Biol. 124 (1-2), 101-115 (1994)

FEATURES	source	Location/Qualifiers
		1..660
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone_lib="Human pheochromocytoma cDNA library"
gene		11..634
		/gene="rab8"
CDS		11..634
		/gene="rab8"
		/codon_start=1
		/product="rab8 small GTP binding protein"
		/protein_id="CAA04065.1"
		/db_xref="GI:452318"
		/db_xref="SWISS-PROT:P24407"
		/translation="WAKTYDYLFKLLLLICDGGVGKTCVLFPRSEDAFNFTFTSTIGID FKRIELDGRKILQWDTAGCRPRTITAYIRGAMGIMLYIDITNKSFDNIRNK IRNIIEHASADVEKIMLGNDVNRQVSRGEKALDYGIKFMETSAKANINVEN AFFTLRIDIKAMDKKLEGNPSQGNQVKITPDQQRKRSFFRCVLL"
BASE COUNT	197 a	159 c 172 g 132 t

Alignment Scores:	
Pred. No.:	3.64e-45
Score:	545.00
Percent Similarity:	77.01%
Best Local Similarity:	52.94%
Query Match:	49.32%
DB:	9
	9
	0
Length:	660
Matches:	99
Conservative:	45
Mismatches:	43
Indels:	0
Gaps:	0

US-09-817-198A-2 (1-212) x HSMRAB8 (1-660)

QY		1	MetAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuIleClyaspSerGlyValGly	20
Db		11	ATGGCGAAGACCTCATGATTACCTGTTCAAAGCTGCTGATCGGGACATCGGGGTGGG	70
QY		21	LysThrCysLeuCysArgPheThrAspAsnGluPheHisSerSerHisIleSerThr	40
Db		71	AAGACCTGTGCTGTTCGGTCTTCGAGGAGCGCTCAACTCCACTTTTATCTCCACC	130

41 IIEGIVAAASPENLYSMETLYSTRNRIIEGLUVAIASPGIILELYSVALRGIIELIN 60
 QY
 Db
 131 ATAGGAATTGACTTTAAAATTAGGCACCATAGAGCTCGATGGCAGAGAAATTAACATGCAG 190
 QY
 61 IIEtPAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLyscInIleTyrArgArg 80
 QY

QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluAcrGSerTyrClnHisIleMet 100
 ||| ||||| :::||| ::|||
Db 251 GCAATGGGCATCATGCTGGTGTCACGACATACCACAGAGAAGTCCTTCGACAACATCCCG 310

[illegible]

Qy	141	LysGluTyrGlyMetAspPheTyrGluThrSerAlaCystThrAsnLeuAasnIleLysGlu	160
	:	: :: :::	: : : : :
Dd	431	CTCGACTATGGAAATCAAGTTTCATGGACACCAGCCGAAGGCCAACATCATGTGGAANAAT	490
Qy	161	SerPhethrArgLeuThrGluLeuValLeuClnAlaHisArgLysGluLeuGluClyleu	180
	:	: :: :::	: : : : :
Dd	491	GCAITTTTTCACTCTCGGCAGAGATATCAAAGCAAAAATTGGACAAAAAATTGGGAAGCCAAC	550
Qy	181	ArgMetArgAlaSerAsnGlu	187
	:	: :: :::	: : : : :
Dd	551	AGCCCCCAGGGGAGCAACCCAG	571
 RESULT 9 CFRAB8 760 bp mRNA linear MAM 30-SEP-1999			
LOCUS			
DEFINITION			
ACCESSION X56385			
VERSION X56385.1 GI:920			
KEYWORDS GTP-binding protein; protein transport; ras gene; ras gene family.			
SOURCE Canis familiaris.			
ORGANISM Canis familiaris			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
REFERENCE	1 (bases 1 to 760)		
AUTHORS	Chavrier,P., Vingron,M., Sander,C., Simons,K. and Zerial,M.		
TITLE	Molecular cloning of YPT1/SEC4-related cDNAS from an epithelial cell line		
JOURNAL	Mol. Cell. Biol.	10 (12),	6578-6585 (1990)
MEDLINE	91061765		
PUBMED	2123294		
REFERENCE	2 (bases 1 to 760)		
AUTHORS	Zerial,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-AUG-1990) Zerial M., EMBL, Meverhofstrasse 1, 6900		

FEATURES	Location/Qualifiers
source	1..760
Heidelberg, FMS	

```

source
1: 700
/organism="Canis familiaris"
/strain="cocker spaniel"
/db_xref="taxon:9615"
/clone="2"
/cell_line="MDCK"

```

```
gene
10.633
/cell_line="MDCK"
cns
10.633
/gene="rab8"
```

```

/genes="rab8"
/function="GTP binding protein"
/codon_start=1
/protein_id="CAB56776.1"
/db_xref="GI:6006436"
/translation="MAKTYDYLFKLLIGDSGVGKTCVLPFSEDAPNFTFTIGID
FKRLTELDGKRILQIWDQAQERFTITAYRGAMGIMLYDITNKSFDINRW
IRNIEHASADVEKMTLGNKVDNRQVSKERGEKALDYGIKFMETSAKANINVE
AFETLHDKAKMDKYLEGNSQGSQGVKITPDQOKRSFRCVLL"
215 a 190 c 196 g 159 t
BASE COUNT
ORIGIN

```

Alignment Scores:	
Pred. No.:	4.32e-45
Score:	545.00
Matches:	99
Conservative:	45
Mismatches:	43
Indels:	0
Gaps:	0
DB:	4
Query Match:	49.32%
Best Local Similarity:	52.94%
Percent Similarity:	77.01%
Score:	545.00
Pred. No.:	4.32e-45
Length: 760	

US-09-817-198A-2 (1-212) x CFR88 (1-760)

```
QY 1 MetalAlysGlnTyrAspValLeuPheArgueuLeuLeuLeuValcly 20
      |||||
Db 10 ATGCGAAGACCTACGATTACCTGTTCAAGCTGCTGCTATCGGGAGCTCGGGGGTGGG 69
      |||||
QY 21 IvstbTrCvstLeuLeuCysArgPheThrAspAsnGlnPheHissSerSerHisTleSerThr 40
```



```
Db 70 AAGACTGTGCTGCTGCTTCCGAGGACGCTTCAACATCAACTTTCCTCCACT 129
QY 41 ILeGlyValAspPheLysMetLysThrIleGluValAspGlyLysValArgIleGln 60
Db 130 ATAGGAATGACTTTAAATATAGACCATAGACTCGATGGCAAGAAATTAAGCTACAG 189
QY 61 ILeTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
Db 190 ATATGGGACACAGCTGGTCAAGAACGGTTTCGGACGATCACAACAGCCTATTACAGGGC 249
QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
Db 250 GCAATGGGCATCATGCTGTCTATGATCATCACAACAGCAAAATTCCTTGACAATATCCGG 309
QY 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
Db 310 AACTGGATTCGGACATTCAGGAGCATGCTTCGAGATGTCGAAAGATGATCTCGGA 369
QY 121 AsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnLeuAla 140
Db 370 AACAAAGTGTGATGTGAACGACAAAGACAAAGTTTCCAAGGACGGGAGAAAAGCTGGCC 429
QY 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
Db 430 CTGCACTATGGAATCAAGTTTCATGAGACCATGTCGAAGGCCAACATCAATGTGGAGAAC 489
QY 161 SerPheThrArgLeuThrGluLeuValGlnAlaHisArgLysGluLeuGluGlyLeu 180
Db 490 GCATTTTCTACTCTCCGAGACATCAAGCAAGCAAGATGGACAAAATTTGGAAGGCAAC 549
QY 181 ArgMetArgAlaSerAsnGlu 187
Db 550 AGTCCCAAGGGAGCAACCAAG 570

RESULT 10
BC002977 2048 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, mel transforming oncogene (derived from cell line
DEFINITION NK14)- RAB8 homolog, clone MGC:2196 IMAGE:3547214, mRNA, complete
cds.
ACCESSION BC002977
VERSION BC002977.1 GI:12804236
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2048)
Strausberg,R.
Direct Submission
Submitted (06-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs@remail.nih.gov
Tissue procurement: ATCC/DCTD/DTP
CDNA Library Preparation: Rubin Laboratory
DNA sequencing by: The I.M.A.G.E. Consortium (LLNL)
BC Cancer Agency, Vancouver, BC, Canada
info@bcqsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Len Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasia van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 8 Row: d Column: 22
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 452317.

FEATURES

source

1..2048
/organism="Homo sapiens"
/db_xref="LocusID:4218"
/db_xref="taxon:9606"
/clone="MGC:2196 IMAGE:3547214"
/tissue_type="Skin, melanotic melanoma."
/clone_lib="NIH_MGC_20"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
77..700
/codon_start=1
/product="mel transforming oncogene (derived from cell line NK14)- RAB8 homolog"
/protein_id="AAH02977.1"
/db_xref="GI:12804237"
/translation="MAKTYDYLFKLLLLIGDSGVGKTVLRFSEDAFNSTFTSTIGIDFKIRTELDGKRIKLIQINDTAGOERFTITAYYRGAMGIMLVYDITNKSNDINRWIKNIEHASADVEKMLGNKNDYQNRQVSKERGEKALDYGIKFMETSAKANINVENAFPTLARDIKAKMDKKLEGNPQSGNQVKITPDQOKRSSFFRCVLL"
BASE COUNT 532 a 556 c 493 g 467 t
ORIGIN

Alignment Scores:

Pred. No.: 1,45e-44 Length: 2048
Score: 545.00 Matches: 99
Percent Similarity: 77.01% Conservative: 45
Best Local Similarity: 52.94% Mismatches: 43
Query Match: 49.32% Indels: 0
DB: 9 Gaps: 0

US-09-817-198a-2 (1-212) x BC002977 (1-2048)

QY 1 MetAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerGlyValGly 20
Db 77 ATGCGAGACCTTACGATTTACCTGTTCAAGCTGCTGATCGGGGACTCGGGGGTGGG 136
QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisIleSerThr 40
Db 137 AAGACCTGTGCTCCTTCCGCTTCTCCGAGGACGCTTCAACTCCACTTTATCTCCACC 196
QY 41 ILeGlyValAspPheLysMetLysThrIleGluValAspGlyLysValArgIleGln 60
Db 197 ATAGGAATTCACCTTTAAATATGAGACCATAGACTCGATGGCAAGAAATTAAGCTCAG 256
QY 61 ILeTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
Db 257 ATATGGGACACAGCGGTCAGAACGGTTTCGGACGATCACAACGGGCTACTACAGGGT 316
QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
Db 317 GCAATGGGCATCATGCTGTCTATGATCATCACAACAGCAAAATTCCTTGACAATCCGG 376
QY 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
Db 377 AACTGGATTCGGACATTCAGGAGCAGCCTCTGCGACACCTCGAAAAGATGATCTCGGG 436
QY 121 AsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnLeuAla 140
Db 437 AACAAAGTGTGATGTGAACGACAAAGATTTCCAAGGACGGGAGAAAAGCTGGCC 496
QY 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
Db 497 CTCGACTATGGAATCAAGTTTCATGAGACCATGTCGAAGGCCAACATCAATGTGGAAAAT 556
QY 161 SerPheThrArgLeuThrGluLeuValGlnAlaHisArgLysGluLeuGluGlyLeu 180
Db 557 GCATTTTCTACTCTCCGAGACATCAAGCAAGCAAGATGGACAAAATTTGGAAGGCAAC 616

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

```

/tissue_type="Liver, normal, 5 month old male mouse."
/clone_lib="NCI_CGAP_Li9"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1. .624
/codon_start=1
/product="Similar to mel transforming oncogene (derived
from cell line NK14)- RAB8 homolog"
/protein_id="AAH19990.1"
/db_xref="GI:18043409"
/translation="MAKTVDFLKLIIIGDSGVKTCVLFRESEDAFNSTFTSTIGID
FKRTTELQGRKLIKLIWDTAGGERFTTITAYRGAMGLMVLITNKRSPFNIRNW
IRNIEHSAADVEKMLGNCKNDKRRQSGEKGLADYGIFKEFTSAKANINVEN
AFPTFLARDJAKMKDKLENGSPQSGSHGKVIITVEQKRTSFPRCSLL"
367 a 341 q 291 t
BASE COUNT

```

Alignment Scores:	
Pred. No.:	2.15e-44
Score:	541.00
Percent Similarity:	75.13%
Best Local Similarity:	50.25%
Query Match:	48.96%
DB:	
DB:	10
Gaps:	0
Indels:	0
Mismatches:	49
Conservative:	49
Matches:	99
Length:	1337

US-09-817-198A-2 (1-212) x BC019990 (1-1337)

QY		1	MetalalysGlnTyrAspValLeupheArgLeuLeuLeuileclysaspSerGlyValgLy	20
Db		1	ATGGCAAGACCTACGATTACCTGTTCAAAGCTGCTGATCGGGACTCGGGGGTAGG	60
QY		21	LysThrCysleuLeuCysArgPheThrAspAsnGluPheHisSerSerHisIleSerThr	40
Db		61	AAGACCTGTCTCTGTCCGCTTCTCCGAGGAGCGCTTCAACTCCACATCATCTACC	120
QY		41	IleGlyValAspPhelysMetLysThrIleGluValAspGlyIleLysValArgIleGln	60
Db		121	ATAGGAATTGCATTAAAATTAGGACCATAGAGCTCGATGGCAAGAGGATTAACTGCAG	180
QY		61	IleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyArgArg	80
Db		181	ATATGGACACGCCCGCCAGGAGCGGTTTCGAACAATCACACACGCTTACAGGGGT	240

QY	81	AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet	100
DB	241	GCCATGGGTATCATGTGCTACGACATACCAATGAGAAGTCCTTTGACAACATCCGG	300
QY	101	LysTrpValSerAspValaspGluTyrAlaProGluGlyValClnLysIleLeuIleGly	120
DB	301	AATGTGATCGGAACATTGAAGAGCATGCTCTGCAGACGTGAGAAAGATGATACTGGG	360
QY	121	AsnLYSAlaaspGluGluGlnLysArgGlnValGlyArgGluClnGlnGlyGlnGlnLeuAla	140
DB	361	AATAAGTGATGTGAATGACAAGAGACAGGTGTCGAAGNAACGGGGAGAAAGCTGGCA	420
QY	141	LysGluTyrGlyMetaspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu	160
DB	421	CTCGACTATGGGATCAAGTTTCATGGAGACACAGTGTCAAAGCCCAACATTATGTGGAGAA	480
QY	161	SerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGlyLeu	180
DB	481	GCATTTTTTCCTCTCCAGGGATATCAAGCAAAAATGGACAAAAAATGGAGGGAAC	540
QY	181	ArgMetArgAlaSerAsnGluLeuAlaLeuAlaGluLeuGluGluGluGlu	197
DB	541	AGCCCGCAGGGGAGCAGCATGGAGTCAAGATCAAGTGCAGTGCAGCAGCAAG	591

RESULT 13	
DYGORAZ2	
LOCUS	DYGORAZ2 765 bp mRNA linear VRT 28-APR-1993
DEFINITION	Discozyme ommata GFP-binding protein (ora2) mRNA, complete cds.
ACCESSION	M38391
VERSION	M38391.1 GI:213118

2

GTP-binding protein.
 D.-omata electric lobe electromotor neuron, cDNA to mRNA.
 Discopage omata
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 Elasmobranchii; Squalea; Hypnosqualea; Pristiorajae; Batoidae;
 Torpediniformes; Narcinoidae; Narcinidae; Discopage.
 1 (bases 1 to 765)
 Ngsee, J.K., Elferink, L.A. and Scheller, R.H.
 A family of ras-like GTP-binding proteins expressed in electromotor
 neurons
 Unpublished (1990)
 Draft entry and computer-readable sequence for [1] kindly submitted
 by J.K.Ngsee, 10-SEP-1990.
 Dept of Biological Sciences
 Herrin Lab
 Stanford University
 Stanford, CA 94305-5020
 Draft entry and computer-readable sequence for [1] kindly submitted
 by J.K.Ngsee, 10-SEP-1990.
 Dept of Biological Sciences
 Herrin Lab
 Stanford University
 Stanford, CA 94305-5020.

FEATURES source

gen

CDS

BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:

US-09-817-198A-2 (1-212) x DYGORA2 (1-765)

QY	20	1	MetAla	lysGlnTyr	aspValLeu	pheArgLeu	leuLeu	leuGly	aspSerGly	valGly	20			
Db	73	14	ATGGC	GAAGACG	TACGATTATCTT	CAAGCTGCTGAT	CGTGATCGGTGAC	AGCGGCGTCGC			73			
QY	40	21	LysThr	CysLeu	LeuCys	ArgPhe	ThrAsp	AsnGlu	pheHis	SerSerHis	LeuSerThr	40		
Db	133	74	AAAAC	TGCCTCTCT	TCCGCTTCTCGG	AGGCGCTTCAAC	ACCACCTTCA	TCTCATCTCCACC			133			
QY	60	41	IleGly	valAsp	PheLys	MetLys	ThrIle	Glu	valAsp	GlyIle	LysValArg	IleGln	60	
Db	193	134	ATCGGA	ATTGATTTC	AAAAATACCA	ACAGTAGA	AGTAAGTATAG	TGGGAAGAA	ATAAAGTCCAG			193		
QY	80	61	IleTrp	AspThr	Ala	GlyGln	GluArg	TyrGln	ThrIle	ThrLys	GlnTyr	TyrArg	Arg	80
Db	253	194	ATATGG	ACACACG	AGGTCA	AGAAAGATTCCG	AACGATACA	ACCGCATCTAC	CCGAGGA				253	
QY	100	81	AlaGln	GlyIle	Phe	LeuVal	TyrAsp	IleSer	SerGlu	ArgSer	TyrGln	HisIle	Met	100
Db	313	254	GCTAT	CGGAATAAT	GAGGTATATG	TACATAACCA	ATGAGTAAGTCTT	TGACACAT	ATAAG				313	

```

101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysLeuLeuGly 120
    |||...:|||||...:|||||
314 AACTGGATAAGGAATATTCAAGAGCATCTTCAGATGTAGAAAGATGATCTGGGC 373
    |||...:|||||...:|||||

121 AsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnLeuAla 140
    |||...:|||||...:|||||
374 ACAAATGGGAGCATGAATGAAAGGCAAGTGTCAAGAGCAAGGAGAAAGTATGCG 433
    |||...:|||||...:|||||

141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnLysGlu 160
    |||...:|||||...:|||||
434 ATAGATTATGGATAAGTCTCTGGAACCCAGCGCAAAATCCAGCATAAACGTTGAAG 493
    |||...:|||||...:|||||

161 SerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGlyLeu 180
    |||...:|||||...:|||||
494 GCATTTATACACTTCGACGAGATATCATGACGAACTCAACAAGAAATG----- 544
    |||...:|||||...:|||||

181 ArgMetArgAlaSerAsnGluLeuAlaLeuAlaGlu---LeuGluGluGluGlyLys 199
    |||...:|||||...:|||||
545 -----AATGAAACAGCTCTCCAGGAGCGACAAAGTAAATCAACCG 589
    |||...:|||||...:|||||

200 ProGluGlyProAlaAsnSerSerLys 208
    |||...:|||||...:|||||
590 CCAAGAACCCAGTCAAAGAAAG 616

RESULT 14
BC000799
LOCUS
DEFINITION Homo sapiens, RAB13, member RAS oncogene family, clone MGC:5074
IMAGE:3451945, mRNA, complete cds.
ACCESSION BC000799
VERSION BC000799.1 GI:12654002
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1211)
Strausberg,R.
Direct Submission
Submitted (15-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAP Plate: 4 Row: 1 Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4506362.
Location/Qualifiers
1. 1211
/organism="Homo sapiens"
/db_xref="Locusid:5872"
/db_xref="taxon:9606"
/clone="MGC:5074 IMAGE:3451945"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NTH_MGC_10"
/lab_host="DH10B"
/Note="Vector: pCMV-SPORT6"
95. .706
CDS

```

```

/codon_start=1
/product="RAB13, member RAS oncogene family"
/protein_id="AAH00799.1"
/db_xref="GI:12654003"
/translation="MAKAYDHLFLKLLLLGDSGVGKTCILIRFAEDFNNTYISTIGID
FKIRTVDIGKKIKLQVWDTAGQRFKTTTAYIRGAMGIIIVYDITDKSENIQNW
MKSCKENASAGVERLLKDCMEARKVKQEQADKLAREHGIRFEFFETSAKSSMNVDE
AFSSLIARDILLKSGRRSGNGKNKPPSTDLKTCDKKNTKCSLG"
BASE COUNT 392 a 256 c 335 g 228 t
ORIGIN
Alignment Scores:
Pred. No.: 8.38e-44 Length: 1211
Score: 534.50 Matches: 102
Percent Similarity: 66.38% Conservative: 54
Best Local Similarity: 43.40% Mismatches: 55
Query Match: 48.37% Indels: 25
DB: 9 Gaps: 1
US-09-817-198A-2 (1-212) x BC000799 (1-1211)
QY 1 MetAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuGlyAspSerGlyValGly 20
    |||...:|||||...:|||||
DB 95 ATGGCCAAAGCCCTACGACCACCTCTTCAAGTTGCTGCTGATCGGGGACTCGGGGGTGGC 154
    |||...:|||||...:|||||
QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisLeSerThr 40
    |||...:|||||...:|||||
DB 155 AGACTGTGTCATCATTCATTCGCTTCGAGAGGACAACTTCAACAACACTTACATCTCCACC 214
    |||...:|||||...:|||||
QY 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyLysValArgLysGln 60
    |||...:|||||...:|||||
DB 215 ATCGGAATTTGATTTCAAGATCGGCACTGTGGATATAGAGGAAAGATCAAACTACAA 274
    |||...:|||||...:|||||
QY 61 IleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTrpArg 80
    |||...:|||||...:|||||
DB 275 GTCTGGGACACGGCTGGCCAGAGCGGTTCAGACAACTAATCACTACCTACCTACCGTGA 334
    |||...:|||||...:|||||
QY 81 AlaGlnGlyIlePheLeuValTyrAspLysSerSerGluArgSerTyrGlnHisIleMet 100
    |||...:|||||...:|||||
DB 335 GCATCGGGCATTTATCTTAGTATACGACATCAGGATGAGAAATCTTTCAGAAATATTCAG 394
    |||...:|||||...:|||||
QY 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysLeuLeuGly 120
    |||...:|||||...:|||||
DB 395 AACTGGATGAAAGCATCAAGGAGATGCTCGGCTGGGTGGGCGCTCTTCTGCTGGG 454
    |||...:|||||...:|||||
QY 121 AsnLysAlaAspGluGlnGluArgGlnValGlyArgGluGlnGlnLeuAla 140
    |||...:|||||...:|||||
DB 455 AACAAATGTGACATGAGGCCAAGAGAGAGGTGCAAGAGGACGAGCCGATAGTTGGCT 514
    |||...:|||||...:|||||
QY 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnLysGlu 160
    |||...:|||||...:|||||
DB 515 CGAGAGCATGGAATCGGATTTTCGAACTAGTCTAAATCCAGTATGATGTGGATGAG 574
    |||...:|||||...:|||||
QY 161 SerPheThrArgLeuThrGluLeuValLeu----- 170
    |||...:|||||...:|||||
DB 575 GCTTTTAGTTCCTGCGCCGGGACATCTTCTCAAGTCAGAGCGCGGAGATCAGGAAAC 634
    |||...:|||||...:|||||
QY 171 -----GlnAlaHisArgLysGlu 176
    |||...:|||||...:|||||
DB 635 GGCACAAAGCCCTCCAGTACTGACCTGAAAATTTGTGACAAGAAAGAA-CACCAACAAGTG 693
    |||...:|||||...:|||||
QY 177 LeuGluGlyLeuArgMetArgAlaSerAsnGluLeuAlaLeuAlaGluLeuGluGlu 196
    |||...:|||||...:|||||
DB 694 CTCCTCGGCTGAGGACCCCTTCTTCCTCCACCACCCCGGAGCTGAACCTGAGGGAGAC 753
    |||...:|||||...:|||||
QY 197 GluGlyLysProGluGlyProAlaAsnSerSerLysThrCysTrp 211
    |||...:|||||...:|||||
DB 754 AACGGCAGAGGAGTGCAGCAGGGGAGAAATAGCAGAGGGGCTTG 798
    |||...:|||||...:|||||
RESULT 15
HSRAB13
LOCUS
DEFINITION H.sapiens mRNA for rab 13.
1238 bp mRNA linear PRI 02-FEB-1994

```

Db	440	ACTGGATGAAAAGCATCAAGAGAAATGCCTCGGCTGGGTGAGCGCCCTCTTCGTGGGG	499
Qy	121	AsnLysAlaaspGluGluGlnLysArgGlnValcylArgGluGlnGlyGlnGlnLeuAla	140
Db	500	AACAAATGTGACATGGAGGCCAAGAGGAAGTGCAGAAAGCAGCAGCCGATAGTGGCT	559
Qy	141	LysGluTyrGlyMetaspPheTyrGluThrSerAlaCysThrAsnLeuAsnLysLysGlu	160
Db	560	CGAGAGCATGGAATCCGATTTTTCGAACCTAGTGTCTAAATCCAGTATGAATGTGGATGAG	619
Qy	161	SerPheThrArgLeuThrGluLeuValLeu-	170
Db	620	GCYTTTAGTTCCTGGCCCGGACACTGCTCAAGTCAGGAGCGCGGAGATCAGGAAC	679
Qy	171	-----GlnAlaHisArgLysGlu	176
Db	680	GGCAACAAGCCTCCAGTACTGACCTGAAACTTGTGACAAGAAGAA-CACCAACAAGTG	738
Qy	177	LeuGluGlyArgMetArgAlaSerAsnGluLeuAlaLeuAlaGluLeuGluGlu	196
Db	739	CTCCTGGGCTGAGGACCCCTTCTTGCCTCCCAACCCCGGAAGCTGAACCTGAGGGAGAC	798
Qy	197	GluGlyLysProGluGlyProAlaAsnSerSerLysThrCysTrp	211
Db	799	AACGGCAGGGAGTGCAGAGGGAGAGAAATAGCAGAGGGGCTTGG	843

Search completed: November 17, 2002, 21:32:05
Job time : 2670 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2002, 18:07:04 ; Search time 67 Seconds
(without alignments)
421.629 Million cell updates/sec

Title: US-09-817-198A-2

Perfect score: 1105

Sequence: 1 MAKQYDVLFRLLLLIGDSGVG.....LEEEGKPGPANSSKTCWC 212

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1105	100.0	401	AAU17136	Novel signal trans
2	1092	98.8	218	AAU17136	Human ORF1368
3	832	75.3	188	AAU17555	Novel signal trans
4	546.5	49.5	204	AB070670	Drosophila melanog
5	545	49.3	213	AB070670	Human rab8 homolog
6	540	48.9	221	AB070670	Human ovarian anti
7	532	48.1	246	AB070670	Lung cancer associ
8	530.5	48.0	200	AA19165	Amino acid sequenc
9	530.5	48.0	200	AA09979	Human Rab10 protei
10	530.5	48.0	200	AA095340	Human protein sequ

11	530.5	48.0	218	21	AA056993	Human prostate can
12	529.5	47.9	218	22	AB07266	Novel human diagno
13	528.5	47.8	207	22	AB071647	Drosophila melanog
14	528.5	47.8	207	22	AB071647	Amino acid sequenc
15	528.5	47.8	207	22	AB071647	Human protein sequ
16	526	47.6	201	21	AA09982	Canine Rab10 prote
17	523	47.3	199	21	AA09980	Human Rab10 protei
18	523	47.3	199	21	AA09981	Human Rab10 protei
19	519	47.0	216	21	AA09868	Arabidopsis thalia
20	519	47.0	216	21	AA09868	Arabidopsis thalia
21	519	47.0	253	21	AA09868	Arabidopsis thalia
22	519	47.0	254	21	AA09868	Arabidopsis thalia
23	513.5	46.5	215	21	AA09868	Zea mays protein f
24	510	46.2	218	21	AA09868	Arabidopsis thalia
25	507	45.9	216	21	AA09868	Arabidopsis thalia
26	506.5	45.8	209	22	AB023365	Novel human diagno
27	501	45.3	224	21	AA07826	Arabidopsis thalia
28	501	45.3	234	21	AA07826	Arabidopsis thalia
29	501	45.3	335	22	AB023366	Novel human diagno
30	478.5	43.3	201	20	AA00919	Human Rab protein,
31	478.5	43.3	201	22	AA028024	Novel human secret
32	478.5	43.3	221	22	AA013525	Human polypeptide
33	478.5	43.3	224	21	AA058758	Breast and ovarian
34	476	43.1	203	21	AA030499	Arabidopsis thalia
35	476	43.1	258	21	AA030498	Arabidopsis thalia
36	468	42.4	202	21	AA07763	Gene 44 human secr
37	467	42.3	205	21	AA034844	Human secreted pro
38	467	42.3	205	21	AA034844	Arabidopsis thalia
39	466	42.2	202	21	AA010858	Arabidopsis thalia
40	462	41.8	205	22	AB059808	Drosophila melanog
41	457	41.4	218	21	AA030710	Arabidopsis thalia
42	457	41.4	221	21	AA030709	Arabidopsis thalia
43	457	41.4	254	21	AA053323	Arabidopsis thalia
44	456.5	41.3	227	21	AA019163	Amino acid sequenc
45	456.5	41.3	227	22	AA039600	Human polypeptide

ALIGNMENTS

RESULT 1
AAU17136
ID AAU17136 standard; Protein; 401 AA.
XX AC AAU17136;
XX AC AAU17136;
DT 07-NOV-2001 (first entry)
XX DE Novel signal transduction pathway protein, Seq ID 701.
XX DE Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
XX DE antinflammatory; anti-HIV; antibacterial; antinflammatory; cancer;
XX DE immune system disorder; rheumatoid arthritis; inflammatory condition;
XX DE organ transplant rejection; infection; hepatitis C; blood disorder;
XX DE sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
XX DE neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
XX DE chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
XX DE cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
XX DE reproductive system; gastrointestinal; liver disorder; AIDS;
XX DE acquired immune deficiency syndrome.
XX OS Homo sapiens.
XX OS WO200154733-A1.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01312.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.

PR	13-OCT-2000;	2000US-02399935
PR	13-OCT-2000;	2000US-02399937
PR	20-OCT-2000;	2000US-0240960
PR	20-OCT-2000;	2000US-0241221
PR	20-OCT-2000;	2000US-0241785
PR	20-OCT-2000;	2000US-0241786
PR	20-OCT-2000;	2000US-0241787
PR	20-OCT-2000;	2000US-0241808
PR	20-OCT-2000;	2000US-0241809
PR	20-OCT-2000;	2000US-0241826
PR	01-NOV-2000;	2000US-0244637
PR	08-NOV-2000;	2000US-0246474
PR	08-NOV-2000;	2000US-0246475
PR	08-NOV-2000;	2000US-0246476
PR	08-NOV-2000;	2000US-0246477
PR	08-NOV-2000;	2000US-0246478
PR	08-NOV-2000;	2000US-0246478
PR	08-NOV-2000;	2000US-0246523
PR	08-NOV-2000;	2000US-0246524
PR	08-NOV-2000;	2000US-0246526
PR	08-NOV-2000;	2000US-0246527
PR	08-NOV-2000;	2000US-0246528
PR	08-NOV-2000;	2000US-0246532
PR	08-NOV-2000;	2000US-0246609
PR	08-NOV-2000;	2000US-0246610
PR	08-NOV-2000;	2000US-0246611
PR	08-NOV-2000;	2000US-0246613
PR	17-NOV-2000;	2000US-0249210
PR	17-NOV-2000;	2000US-0249211
PR	17-NOV-2000;	2000US-0249212
PR	17-NOV-2000;	2000US-0249213
PR	17-NOV-2000;	2000US-0249214
PR	17-NOV-2000;	2000US-0249215
PR	17-NOV-2000;	2000US-0249216
PR	17-NOV-2000;	2000US-0249217
PR	17-NOV-2000;	2000US-0249218
PR	17-NOV-2000;	2000US-0249244
PR	17-NOV-2000;	2000US-0249245
PR	17-NOV-2000;	2000US-0249264
PR	17-NOV-2000;	2000US-0249265
PR	17-NOV-2000;	2000US-0249267
PR	17-NOV-2000;	2000US-0249299
PR	17-NOV-2000;	2000US-0249300
PR	01-DEC-2000;	2000US-0250160
PR	01-DEC-2000;	2000US-0250391
PR	05-DEC-2000;	2000US-0251030
PR	05-DEC-2000;	2000US-0251031
PR	05-DEC-2000;	2000US-0251988
PR	05-DEC-2000;	2000US-0256719
PR	06-DEC-2000;	2000US-0251479
PR	08-DEC-2000;	2000US-0251856
PR	08-DEC-2000;	2000US-0251868
PR	08-DEC-2000;	2000US-0251869
PR	08-DEC-2000;	2000US-0251989
PR	08-DEC-2000;	2000US-0251990
PR	11-DEC-2000;	2000US-0254037
PR	05-JAN-2001;	2001US-0259678
XX	(HUMA-) HUMAN GENOME SCI INC	
XX		
PI	Rosen CA, Barash SC, Ruben	
XX	WPI; 2001-465460/50.	
DR	N-PSDB; AAS27053.	
XX		
XX	Novel polypeptides useful for	
PT	prognosing disorders related	
PT	to Alzheimer's disease, mental disorders and neuronal disorders	
XX		
PT	Claim 1; SEQ ID No 701; 8809pt	
XX		

Claim 1; SEQ ID No 701; 880pp; English.

CC The invention relates to novel isolated polypeptides (I), and
 CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
 CC diagnosing, preventing and treating diseases including immune system
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
 CC transplant rejections and graft versus host disease, infectious diseases
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
 CC other blood-related disorders (sickle cell anaemia), myeloproliferative
 CC disorders, primary haematopoietic disorders, hyperproliferative
 CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
 CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders
 CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
 CC Addison's disease), reproductive system disorders, gastrointestinal
 CC disorder (inflammatory disorders), liver disorders (cirrhosis),
 CC as stimulators of B-cell responsiveness to pathogens, activators of
 CC T-cells, to induce higher affinity antibodies, and as a means to induce
 CC tumour proliferation in pathologies e.g. acquired immune deficiency
 CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
 CC pathway protein, amino acid sequences of the invention.
 XX

Query Match 100.0%; Score 1105; DB 22; Length 401;
 Best Local Similarity 100.0%; Pred. No. 2.1e-108;
 Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSHSTIGVDFKMTIEVDGKVRQ 60
 DB 27 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSHSTIGVDFKMTIEVDGKVRQ 86
 QY 61 IWDTAGQERYQITTKQYRRAGIFLVYDISSERSYQHIMKWSVDVEYAPGVQKILIG 120
 DB 87 IWDTAGQERYQITTKQYRRAGIFLVYDISSERSYQHIMKWSVDVEYAPGVQKILIG 146
 QY 121 NKADEQKRVQREGQOQOLAKKEYGMDFYETSACTNLNIKESFRTLTELVLQAHRL 180
 DB 147 NKADEQKRVQREGQOQOLAKKEYGMDFYETSACTNLNIKESFRTLTELVLQAHRL 206
 QY 181 RMRASNELALAELEEKGKPEGPANSSKTCWC 212
 DB 207 RMRASNELALAELEEKGKPEGPANSSKTCWC 238

RESULT 2
 ABA41604
 ID ABA41604 standard; Protein; 218 AA.
 XX
 AC ABA41604;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF1368 polypeptide sequence SEQ ID NO:2736.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.

XX 05-OCT-2000.
 XX 31-MAR-2000; 2000WO-US08621.
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX (CURA-) CURAGEN CORP.
 XX Shinkets RA, Leach M;
 PI WPI; 2000-602362/57.
 XX N-PSDB; RAC75813.
 DR Novel nucleic acids and peptides derived from open reading frame X,
 XX neurodegenerative disorders and cardiovascular disease -
 PT Claim 11; Page 1979-1980; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 218 AA;
 Query Match 98.8%; Score 1092; DB 21; Length 218;
 Best Local Similarity 97.2%; Pred. No. 2e-107;
 Matches 212; Conservative 0; Mismatches 0; Indels 6; Gaps 1;
 QY 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSHSTIGVDFKMTIEVDGKVRQ 60
 DB 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSHSTIGVDFKMTIEVDGKVRQ 60
 QY 61 IWDTAGQERYQITTKQYRRAGIFLVYDISSERSYQHIMKWSVDVEYAPGVQKILIG 120
 DB 61 IWDTAGQERYQITTKQYRRAGIFLVYDISSERSYQHIMKWSVDVEYAPGVQKILIG 120
 QY 121 NKADEQKRVQREGQOQO-----QLAKKEYGMDFYETSACTNLNIKESFRTLTELVLQAH 174
 DB 121 NKADEQKRVQREGQOQOQPSLOLAKKEYGMDFYETSACTNLNIKESFRTLTELVLQAH 180
 QY 175 KELEGLRMRASNELALAELEEKGKPEGPANSSKTCWC 212
 DB 181 KELEGLRMRASNELALAELEEKGKPEGPANSSKTCWC 218
 RESULT 3
 AAU17555
 ID AAU17555 standard; Protein; 188 AA.
 XX
 AC AAU17555;

XX	07-NOV-2001	(first entry)	
DT	Novel signal transduction pathway protein, Seq ID 1120.		
XX			
DE			
XX			
XX			
KW	Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;		
KW	antiinflammatory; anti-HIV; antibacterial; antibacterial; antiinflammatory; cancer;		
KW	immune system disorder; rheumatoid arthritis; inflammatory condition;		
KW	organ transplant rejection; infection; hepatitis C; blood disorder;		
KW	sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;		
KW	neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;		
KW	chromosomal abnormality; Down syndrome; ischaemia; renal disorder;		
KW	cardiovascular; respiratory; wound healing; endocrine; Addison's disease;		
KW	reproductive system; gastrointestinal; liver disorder; AIDS;		
KW	acquired immune deficiency syndrome.		
XX			
OS	Homo sapiens.		
XX			
XX	WO200154733-A1.		
PN			
XX			
PD	02-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US01312.		
XX			
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180528.		
PR	24-FEB-2000; 2000US-0184564.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216847.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225447.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	14-AUG-2000; 2000US-0225759.		
PR	22-AUG-2000; 2000US-0226279.		
PR	22-AUG-2000; 2000US-0226681.		
PR	22-AUG-2000; 2000US-0226868.		
PR	22-AUG-2000; 2000US-0227182.		
PR	23-AUG-2000; 2000US-0227009.		
PR	30-AUG-2000; 2000US-0228924.		
PR	01-SEP-2000; 2000US-0229287.		
PR	01-SEP-2000; 2000US-0229343.		
PR	01-SEP-2000; 2000US-0229344.		
PR	01-SEP-2000; 2000US-0229345.		
PR	05-SEP-2000; 2000US-0229509.		
PR	05-SEP-2000; 2000US-0229513.		
PR	06-SEP-2000; 2000US-0230437.		
PR	06-SEP-2000; 2000US-0230438.		
PR	08-SEP-2000; 2000US-0231242.		
PR	08-SEP-2000; 2000US-0231243.		
PR	08-SEP-2000; 2000US-0231244.		
PR	08-SEP-2000; 2000US-0231413.		
PR	08-SEP-2000; 2000US-0231414.		
PR	08-SEP-2000; 2000US-0232080.		
PR	08-SEP-2000; 2000US-0232081.		
PR	12-SEP-2000; 2000US-0231968.		
PR	14-SEP-2000; 2000US-0232397.		
PR	14-SEP-2000; 2000US-0232398.		
PR	14-SEP-2000; 2000US-0232399.		
PR	14-SEP-2000; 2000US-0232400.		
PR	14-SEP-2000; 2000US-0232401.		
PR	14-SEP-2000; 2000US-0233063.		
PR	14-SEP-2000; 2000US-0233064.		
PR	14-SEP-2000; 2000US-0233065.		
PR	21-SEP-2000; 2000US-0234223.		
PR	21-SEP-2000; 2000US-0234274.		
PR	25-SEP-2000; 2000US-0234997.		
PR	25-SEP-2000; 2000US-0234998.		
PR	26-SEP-2000; 2000US-0235484.		
PR	27-SEP-2000; 2000US-0235834.		
PR	27-SEP-2000; 2000US-0235836.		
PR	29-SEP-2000; 2000US-0236327.		

PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251989.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-465460/50.
 DR N-PSDB; AAS27472.
 XX
 PT Novel polypeptides useful for diagnosing, treating, preventing and/or
 PT prognosing disorders related to the proteins, including cancers, immune
 PT disorders and neuronal disorders -
 XX
 PS Claim 1; SEQ ID NO 1120; 880pp; English.
 XX
 CC The invention relates to novel isolated polypeptides (I), and
 CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
 CC diagnosing, preventing and treating diseases including immune system
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
 CC transplant rejections and graft versus host disease, infectious diseases
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
 CC other blood-related disorders (sickle cell anaemia), myeloproliferative
 CC disorders, primary haematopoietic disorders, hyperproliferative
 CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
 CC abnormalities (Down syndrome) ischaemic injury (e.g. stroke), renal
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders
 CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
 CC Addison's disease), reproductive system disorders, gastrointestinal
 CC disorder (inflammatory disorders), liver disorders (cirrhosis),
 CC as stimulators of B-cell responsiveness to pathogens, activators of
 CC T-cells, to induce higher affinity antibodies, and as a means to induce
 CC tumour proliferation in pathologies e.g. acquired immune deficiency
 CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
 CC pathway protein, amino acid sequences of the invention.
 XX

AC ABB70670;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 38802.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL14773.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 38802; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AB557737-AB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 204 AA;
 Query Match 49.5%; Score 546.5; DB 22; Length 204;
 Best Local Similarity 49.8%; Pred. NO. 1.4e-49;
 Matches 105; Conservative 47; Mismatches 50; Indels 9; Gaps 3;
 QY 1 MAKQ-YDVLFRLLIGDSGVGKTCCLCRFTDNEFHSHSHISTIGVDFKMKTIYVDGKVR 59
 DB 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSHSHISTIGVDFKMKTIYVDGKVR 60
 QY 60 QWDTAGQERYQTITKQYRRAGIFLVYDISSERSYQHIMKWSDVDEYAPGVQKIL 119
 DB 61 QWDTAGQERYQTITKQYRRAGIFLVYDISSERSYQHIMKWSDVDEYAPGVQKIL 120
 QY 120 GNKADEQKRVGREGQOQLAKEYGDMFYETSACTNLNKESTFRTLTELVLQAHKELEG 179
 DB 121 GNKCDMTKRVYNGERGAIAREHGIRFMETSAKSNINIERAFCELAEAALD----KTSG 176
 QY 180 LMRASNELALAELEEECKPEGPANSSKTC 210
 DB 177 RESAENQERVIIDRRNQEKAP----GYSKCC 203
 RESULT 5
 ABB11916
 ID ABB11916 standard; peptide; 213 AA.
 XX
 AC ABB11916;

XX	11-JAN-2002	(first entry)	
DT			
XX	Human rab8 homologue, SEQ ID NO:2286.		
DE			
XX			
XX	Human; cytokine; cell proliferation; cell differentiation; growth factor;		
KW	haematopoiesis regulation; tissue growth; immunomodulator; activin;		
KW	inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;		
KW	proliferation; metastasis; cancer; tumour; haematopoietic disorder;		
KW	myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;		
KW	chronic inflammatory condition; proliferative retinopathy;		
KW	atherosclerosis; coronary heart disease; arterial ischaemia;		
KW	bone disorder; osteoporosis; vascular growth disorder;		
KW	tissue regeneration; wound healing; infection; immune disorder;		
KW	cell culture; drug screening; gene therapy; antiinflammatory;		
KW	antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;		
KW	cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;		
KW	antifungal; vulnery; antiulcer.		
XX			
XX	Homo sapiens.		
OS			
XX	WO200157188-A2.		
PN			
XX	09-AUG-2001.		
PD			
XX	05-FEB-2001; 2001WO-US03800.		
PF			
XX	03-FEB-2000; 2000US-0496914.		
PR			
XX	27-APR-2000; 2000US-0560875.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
XX	Tang YT, Liu C, Drmanac RT;		
PI			
XX	WPI; 2001-457740/49.		
DR	N-PSDB; ABA09160.		
XX			
XX	Human proteins and DNA encoding sequences useful for preventing,		
PT	treating or ameliorating a medical condition in a mammalian subject		
PT	e.g. arthritis and cancer -		
XX			
PS	Claim 20; Page 276; 1963pp; English.		
XX			
XX	Sequences ABB10981-ABBI2330 represent 1350 novel human polypeptides, and		
CC	sequences ABA08225-ABA09574 represent nucleic acids encoding them. The		
CC	invention also relates to vectors and recombinant host cells comprising a		
CC	nucleotide of the invention, methods of producing the novel polypeptides,		
CC	antibodies against the polypeptides, methods of detecting the nucleotides		
CC	or polypeptides in a sample, and methods of identifying compounds which		
CC	bind to polypeptides of the invention. Although novel, many of the		
CC	polypeptides of the invention have homology to known proteins, thereby		
CC	giving an insight into their probable biological activities, and hence		
CC	potential therapeutic applications. The polypeptides of the invention may		
CC	have various activities, including cytokine, cell proliferation or cell		
CC	differentiation activities; stem cell growth factor activity;		
CC	haematopoiesis regulatory activity; tissue growth activity;		
CC	immunomodulatory activity; activin- or inhibin-related activities;		
CC	chemotactic or chemokinetic activities; haemostatic, thrombotic or		
CC	thrombolytic activities; receptor or ligand activities; or may be		
CC	involved in oncogenesis, cancer cell proliferation or metastasis.		
CC	Depending on their biological activities, polypeptides and nucleotides of		
CC	the invention are useful for preventing, treating or ameliorating medical		
CC	conditions, e.g. by protein or gene therapy. Such conditions include		
CC	cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell		
CC	disorders), chronic inflammatory conditions (e.g., asthma or arthritis),		
CC	proliferative retinopathy, atherosclerosis, coronary heart disease,		
CC	arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal		
CC	vascular growth. Polypeptides involved with tissue regeneration and		
CC	repair (or nucleic acids encoding them) may be used to promote wound		
CC	healing (e.g., of burns, incisions and ulcers), while those with		
CC	immunomodulatory activities may be used in the treatment of viral,		
CC	bacterial and fungal infections in addition to immune disorders.		
CC	Polypeptides with growth factor activity may be used in cell cultures to		

CC		promote cell growth. For example, such polypeptides may be used to
CC		manipulate stem cells in culture to give rise to neuroepithelial cells
CC		that can be used to augment or replace cells damaged by illness,
CC		autoimmune disease or accidental damage. The polypeptides and nucleotides
CC		may also be used in the diagnosis of the above conditions, and in drug
CC		screening techniques. The present sequence represents a novel human
xx		polypeptide of the invention.
xx		
SQ	Sequence	213 AA;
	Query Match	49.3%; Score 545; DB 22; Length 213;
	Best Local Similarity	52.9%; Pred. No. 2.le-49;
	Matches	99; Conservative 45; Mismatches 43; Indels 0; Gaps 0;
Oy	1	MAKQDVLFRLLLIGDSGVGKTCLCRFTDNFHSHTISCTVDFPKMTIEVDGIKVRIQ 60
Dd	7	MAKYDYLFKLLIGDSGVGKTCVLFRESDAFNSTFTSIDFKRTIELDGKRILQ 66
Oy	61	IWDTAGQERYQTITKOYRYRAAGFLVYDVISERSYQHIMKWVSVDVEYAPGVOKILIG 120
Dd	67	IWDTAGQERETITPAYRYGAMGLVYDITNEKSFDNIRNWIRNIEHASADVEKMILG 126
Oy	121	NKADEEQRVRGREGQQOLAKNEYGMDFETSACTNLNKESFTRTELVLQAHRELEGL 180
Dd	127	NKCVDNDXRVSKERGEKLALDYGIKFMEYSAKANINVENAFFTLARDIKAKMDKKLEGN 186
Oy	181	RMRASNE 187
Dd	187	SPQGSNQ 193
RESULT 6		
ABP41333		
ID	ABP41333	standard; Protein; 221 AA.
XX	AC	ABP41333;
XX	AC	
DT	23-AUG-2002	(first entry)
XX	Human ovarian antigen HCGNA67,	SEQ ID NO:2465.
KW	Human; ovarian antigen; ovary; ovarian; breast; cancer; cancer; tumour;	
KW	ovarian cancer; breast cancer; reproductive system disorder;	
KW	infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;	
KW	PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;	
KW	inflammatory condition; immune disorder; blood disorder;	
KW	cardiovascular disorder; respiratory disorder; neurological disorder;	
KW	gastrointestinal disorder; urinary system disorder; drug screening;	
KW	gene therapy; chromosome mapping; forensic analysis;	
KW	antibody preparation; cytostatic; immunomodulatory; neuroprotective;	
KW	antiinflammatory; gynaecological; reproductive.	
XX	Homo sapiens.	
OS		
XX	WO200200677-A1.	
PN		
XX	03-JAN-2002.	
PD		
XX	07-JUN-2001; 2001WO-US18569.	
FF		
XX	07-JUN-2000; 2000US-209467P.	
PR		
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA		
XX	Birse CE, Rosen CA;	
PI		
XX	WPI; 2002-147878/19.	
DR	N-PSDB; ABQ54410.	
DR		
XX		
PT	Isolated nucleic acid molecules encoding novel ovarian polypeptides,	
PT	useful in the prevention, treatment and diagnosis of cancer (e.g.	
PT	ovarian cancer), immune disorders, cardiovascular disorders and	
PT	neurological diseases -	

XX PS Claim 11; SEQ ID No 2465; 2922pp; English.

XX CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ASQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention.
CC Note: The reference data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 221 AA;

Query Match 48.9%; Score 540; DB 23; Length 221;
Best Local Similarity 52.7%; Pred. No. 7.5e-49;
Matches 98; Conservative 45; Mismatches 43; Indels 0; Gaps 0;

Qy 2 AKQYDVLFRLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDPKMTIEVDGKVRQI 61
Db 16 AKTYDYLFRLLIGDSGVGKTCVLFVSEDAFNFTISTIGIDFKIRTVDEGKKIKLI 75

Qy 62 WDTAGQERYOTITTKQYYRRAQGIYFLVDISSERSYQHIMKMWSDVDEYAPGVOKILIGN 121
Db 76 WDTAGQERFRTITTAAYRGAMGIMLVYDITNEKSPDNIRNIRNEEHSADVEKMLIGN 135

Qy 122 KADEQKQVREGQOQOLAKYGMDFYETSACTNLNKESTRTELVLQAHKRELEGLR 181
Db 136 KCDVNDKQVSKERGEKALDYGIKFMETSAKANINVENAFPTLARDIKAKMDKKLEGS 195

Qy 182 MRASNE 187
Db 196 PQGSNQ 201

RESULT 7
AAB58196
ID AAB58196 standard; Protein; 246 AA.

XX AC AAB58196;
XX DT 14-MAR-2001 (first entry)
XX DE Lung cancer associated polypeptide sequence SEQ ID 534.
XX Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease.
XX OS Homo sapiens.

XX WO2000055180-A2.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US05918.
XX PR 12-MAR-1999; 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX Ruben SM;
XX WPI: 2000-587514/55.
XX N-PSDB: AAF18072.
XX Lung cancer associated gene sequences, referred to as lung cancer
XX PT antigens, useful for treatment, prevention, and diagnosis of disorders
XX PT such as lung cancer -
XX PS Claim 11; Page 1027-1028; 1425pp; English.
XX CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
XX CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
XX CC associated proteins and polynucleotide sequences, their agonists, and
XX CC antagonists may have neuroprotective; cytostatic; cardioactive;
XX CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
XX CC general; nephrotropic; antiinfective; gynecological; or antibacterial
XX CC activity. The invention also includes antibodies specific for the
XX CC protein or polynucleotide sequences. The lung cancer associated
XX CC polynucleotide sequences may be used for detection of lung cancer,
XX CC chromosome identification, as chromosome markers, and for numerous other
XX CC diagnostic or research purposes. The proteins may be used to treat
XX CC disorders such as neural, immune, muscular, reproductive,
XX CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX CC disorders. The proteins may also be used in the treatment of wounds and
XX CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
XX CC peptide AAB58549 are used in the course of the invention for the
XX CC identification and characterisation of the polynucleotide and protein
XX CC sequences.

SQ Sequence 246 AA;

Query Match 48.1%; Score 532; DB 21; Length 246;
Best Local Similarity 47.6%; Pred. No. 6.2e-48;
Matches 100; Conservative 49; Mismatches 43; Indels 18; Gaps 3;

Qy 1 MAKQYDVLFRLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDPKMTIEVDGKVRQI 60
Db 44 MAKAYDHLFKLLIGDSGVGKTCLLIRFADNFNNTYISTIGIDFKIRTVDEGKKIKLI 103

Qy 61 IWDTAGQERYOTITTKQYYRRAQGIYFLVDISSERSYQHIMKMWSDVDEYAPGVOKILIG 120
Db 104 VMDTAGQERFRTITTAAYRGAMGIMLVYDITNEKSPDNIRNIRNEEHSADVEKMLIGN 163

Qy 121 NKADBEQKQVREGQOQOLAKYGMDFYETSACTNLNKESTRTELVLQAHKRELEGLR 180
Db 164 NKCDWEAKRKVQEQADKLAREHGIRFFETSAKSMNVDEAFSSLDAIL-----LASK 217

Qy 181 RMRASNELALAELEEEEGKPEGPANSSKTC 210
Db 218 GRRSGN-----GNKP--PSTDKTC 235

RESULT 8
AAB19165
ID AAB19165 standard; Protein; 200 AA.
XX AC AAB19165;
XX DT 19-FEB-2001 (first entry)
XX DT

DE Amino acid sequence of human RAB10.
KW RAB protein; GTPase; GTP binding; gene therapy; cancer.
XX Homo sapiens.
OS
XX
PN WO200058464-A2.
XX
XX 05-OCT-2000.
XX
XX 13-MAR-2000; 2000WO-US06330.
PF
XX 25-MAR-1999; 99US-0126083.
PR
XX
PA (AXYS-) AXYS PHARM INC.
XX
XX Allen M, Abel K, McIntosh B, Vega R, Rutter M, Buckler AJ;
PI
XX WPI; 2000-647233/62.
DR N-PSDB; AAA96887.
XX
XX Novel isolated nucleic acid encoding a mammalian RAB protein useful for
PT identifying homologous or related genes, in producing composition that
PT modulates expression or function of RAB for cancer therapy -
XX
XX Claim 2; Page 39; 58pp; English.
XX
XX The present sequence represents a mammalian RAB protein. RAB proteins
CC constitute the largest family of small GTPases, with over 40 identified
CC isoforms. RAB proteins contain four highly conserved peptide sequences
CC involved in GTP binding and hydrolysis. Compositions comprising RAB
CC nucleic acid are useful for identifying homologous or related genes,
CC in producing compositions that modulate the expression or function of
CC RAB, for gene therapy, mapping functional regions of the protein and
CC in studying associated physiological pathways. In addition, modulation
CC of the gene activity in vivo is used for prophylactic and therapeutic
CC purposes, such as treatment of cancer, and identification of cell type
CC based on expression. The DNA may also be used to identify expression of
CC the gene in a biological specimen.
XX
XX Sequence 200 AA;
XX
Query Match 48.0%; Score 530.5; DB 21; Length 200;
Best Local Similarity 55.8%; Pred. No. 6.6e-48;
Matches 96; Conservative 46; Mismatches 29; Indels 1; Gaps 1;
QY 1 MAKQ-YDVLFRLLIGDGSVGKTCVLFREFSDDAFNTFTSTIGIDFKIKTVELQKKIKL 59
DB 1 MAKTTYDLLFKLLIGDGSVGKTCVLFREFSDDAFNTFTSTIGIDFKIKTVELQKKIKL 60
QY 60 QIWDTAGQERYQTITKQYRRAGIFLVYDISERSYQHIMKWSVDVEYAPGVQKILI 119
DB 61 QIWDTAGQERFHTTTSYRGAMGIMLVYDITNGKSFENISKWLNRNIDSHANEDVERMLL 120
QY 120 GNKADEEQRQVREGOQQLAKYGMDFYETSACTNLNLIKESFTRLTELVLQ 171
DB 121 GNKCDMDKRRVVPKGEQIAREHGIRFFETSAKANINIEKAFLTLAEDILR 172
RESULT 9
AAB09979 standard; Protein; 200 AA.
XX
XX AAB09979;
XX
XX 19-OCT-2000 (first entry)
DT
XX Human Rab10 protein.
DE
XX Human; Rab10; Yap/Rab family.
KW
XX Homo sapiens.
OS
XX (HELI-) HELIX RES INST.

PN CN1249345-A.
XX
PD 05-APR-2000.
XX
PF 28-SEP-1998; 98CN-0121911.
XX
PR 28-SEP-1998; 98CN-0121911.
XX
XX (UYFU-) UNIV FUDAN.
XX
XX Yu L, Tu Q, Gao J;
PI
XX WPI; 2000-400723/35.
DR N-PSDB; AAA40104.
XX
XX Preparation of human gene coding sequence, its encoded polypeptide -
PT
XX Claim 2; Page 18; 23pp; Chinese.
PS
XX
XX This invention describes a novel human Rab10 cDNA sequence. The protein
CC coded by said sequence belongs to Yap/Rab family and is the homolog of
CC mouse Rab10. The present invention also relates to the polypeptide coded
CC by said nucleotide sequence and the application and preparing process
CC of said polynucleotide and said polypeptide. This sequence represents the
CC human Rab10 protein described in the method of the invention.
XX
XX Sequence 200 AA;
XX
Query Match 48.0%; Score 530.5; DB 21; Length 200;
Best Local Similarity 55.8%; Pred. No. 6.6e-48;
Matches 96; Conservative 46; Mismatches 29; Indels 1; Gaps 1;
QY 1 MAKQ-YDVLFRLLIGDGSVGKTCVLFREFSDDAFNTFTSTIGIDFKIKTVELQKKIKL 59
DB 1 MAKTTYDLLFKLLIGDGSVGKTCVLFREFSDDAFNTFTSTIGIDFKIKTVELQKKIKL 60
QY 60 QIWDTAGQERYQTITKQYRRAGIFLVYDISERSYQHIMKWSVDVEYAPGVQKILI 119
DB 61 QIWDTAGQERFHTTTSYRGAMGIMLVYDITNGKSFENISKWLNRNIDSHANEDVERMLL 120
QY 120 GNKADEEQRQVREGOQQLAKYGMDFYETSACTNLNLIKESFTRLTELVLQ 171
DB 121 GNKCDMDKRRVVPKGEQIAREHGIRFFETSAKANINIEKAFLTLAEDILR 172
RESULT 10
AAB95340 standard; Protein; 200 AA.
XX
XX AAB95340;
XX
XX 26-JUN-2001 (first entry)
DT
XX Human protein sequence SEQ ID NO:17619.
DE
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW
XX Homo sapiens.
OS
XX EP1074617-A2.
PN
XX 07-FEB-2001.
PD
XX
XX 28-JUL-2000; 2000EP-0116126.
PF
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.


```
PD 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS71453.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID NO 37625; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 218 AA;
QY Query Match 47.9%; Score 529.5; DB 22; Length 218;
DB Best Local Similarity 55.8%; Pred. No. 9.6e-48;
Matches 96; Conservative 46; Mismatches 29; Indels 1; Gaps 1;
QY 1 MAKQ-YDVLFRLLIGDSGVGKTCILCRFTDNEPHSSHISTIGVDFKMKTIYVDGKIVRI 59
DB 19 MAKRTYDLLEKLLIGDSGVGKTCVLPFRSDDAFNTFTISTIGDFKIKTVELQGGKIKL 78
QY 60 QIWTAGQERYQTITTKOYIRRAQGIFLVYDISSERSYQHIMKWVSDYDEVAPEGVQKILI 119
DB 79 QIWTAGQERPHITTSYRGAMGIMLVYDTNGKSFENISKVLINIDEHANEDVERMLL 138
QY 120 GNKADEQKROVGREGOQQLAKEYGMDFYETSACTNLNIKESFTRLTFLVLQ 171
DB 139 GNKCDMDKRVVPKGKEQIAREHGIRFFETSAKANINIEKAFTLTAEIDILR 190
RESULT 13
ABBY1647
ID ABBY1647 standard; Protein; 207 AA.
XX ABBY1647;
XX ABBY1647;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 41733.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW
```

```
KW pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL15750.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure; SEQ ID NO 41733; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABJ01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 207 AA;
QY Query Match 47.8%; Score 528.5; DB 22; Length 207;
DB Best Local Similarity 53.8%; Pred. No. 1.1e-47;
Matches 100; Conservative 41; Mismatches 40; Indels 5; Gaps 1;
QY 1 MAKQYDVLFRLLIGDSGVGKTCILCRFTDNEPHSSHISTIGVDFKMKTIYVDGKIVRIQ 60
DB 1 MAKTYDYLFRLLIGDSGVGKTCILFRSEDAFNTFTISTIGDFKIKTIELDNKKIKLQ 60
QY 61 IWDTAGQERYQTITTKOYIRRAQGIFLVYDISSERSYQHIMKWVSDYDEVAPEGVQKILIG 120
DB 61 IWDTAGQERFRTITTYRGAMGIMLVYDTQKSFENIKNWIENASADVERKMLLG 120
QY 121 NKADEQKROVGREGOQQLAKEYGMDFYETSACTNLNIKESFTRLTFLVLQAHKKEGL 180
DB 121 NKCELTDKROVSKERGEQIAIEYGIKFMETSAKASINVEAF-----LTLASDIKARTEK 175
QY 181 RMRASN 186
DB 176 RMEANN 181
RESULT 14
AAG67154
ID AAG67154 standard; Protein; 207 AA.
XX AAG67154;
XX AAG67154;
XX 13-NOV-2001 (first entry)
XX Amino acid sequence of human 27423 G-protein.
XX Human; G-protein; 32705; 23224; 27423; 32700; 32712; lung disorder;
XX congenital anomaly; pulmonary congestion; oedema; haemorrhage;
KW
```


Search completed: November 17, 2002, 20:38:26
Job time : 68 secs

Db 241 GCGAGGAGATACCGAGCCATCACAAAGCATACTATCGCGGGGCCCGAGGGGATATTTT 300
QY 301 TGGTCTATGACATTAAGCAGCGCGCTCTTACCAGCACATCATGAAGTGGGTGAGTGAG 360
Db 301 TGGTCTATGACATTAAGCAGCGCGCTCTTACCAGCACATCATGAAGTGGGTGAGTGAG 360
QY 361 TGGATGAGTACGACAGAGCGGTCCAGAGATCCTTATGGGAATAAGGCTGATGAGG 420
Db 361 TGGATGAGTACGACAGAGCGGTCCAGAGATCCTTATGGGAATAAGGCTGATGAGG 420
QY 421 ACCAGAAACCGCAGGTGGGAAGAGCAGAGCGCAGCAGCTGGCAGGAGTATGCGATGG 480
Db 421 AGCAGAAACCGCAGGTGGGAAGAGCAGAGCGCAGCAGCTGGCAGGAGTATGCGATGG 480
QY 481 ACTTCTATGAACAAGTGCCTGCACCAACTCAACATTAAGAGATCATTCACGCGCTCA 540
Db 481 ACTTCTATGAACAAGTGCCTGCACCAACTCAACATTAAGAGTCAATTCACGCGCTCA 540
QY 541 CAGAGCTGGTGTGAGGCCCATAGGAAGAGCTGGAAGGCCCTCCGGATCGCTGCCAGCA 600
Db 541 CAGAGCTGGTGTGAGGCCCATAGGAAGAGCTGGAAGGCCCTCCGGATCGCTGCCAGCA 600
QY 601 ATGAGTTGGCACTGGCAGAGCTGGAGGAGGAGGAGGCAACCCAGAGGCCCAAGCACT 660
Db 601 ATGAGTTGGCACTGGCAGAGCTGGAGGAGGAGGAGGCAACCCAGAGGCCCAAGCACT 660
QY 661 CTTGAAAAACCTGCTGGTGTGAGTCTGTGTGGGCGACCCACACAGCACACCCCTCTTC 720
Db 661 CTTGAAAAACCTGCTGGTGTGAGTCTGTGTGGGCGACCCACACAGCACACCCCTCTTC 720
QY 721 CTCAGAGGCCGTGGGCGAGACAGGGGAGCGGGGCTTTGCCCTGCTGTCTCTCTCGT 780
Db 721 CTCAGAGGCCGTGGGCGAGACAGGGGAGCGGGGCTTTGCCCTGCTGTCTCTCTCGT 780
QY 781 GTGATCACCTATTTAGTATACAGTACCCACTACTGCCCTGCTGTCTCTCTCTCGT 840
Db 781 GTGATCACCTATTTAGTATACAGTACCCACTACTGCCCTGCTGTCTCTCTCTCGT 840
QY 841 CTGCTGTCTATCTCAAGCAGCCCTGTCCCGAGCCCTCCACCCCTGGAGTGGTCTTCTCA 900
Db 841 CTGCTGTCTATCTCAAGCAGCCCTGTCCCGAGCCCTCCACCCCTGGAGTGGTCTTCTCA 900
QY 901 GCCTGTTTCCCGAGCCAGGCGCTGTACGACCCCGCAGATGTGCGCGAAGCACTGTCTC 960
Db 901 GCCTGTTTCCCGAGCCAGGCGCTGTACGACCCCGCAGATGTGCGCGAAGCACTGTCTC 960
QY 961 ACCATCCCGCAGCCACAGACAGCAGCGGCTGAGTGCCAGGCCACTTTTCAGCTGCTC 1020
Db 961 ACCATCCCGCAGCCACAGACAGCAGCGGCTGAGTGCCAGGCCACTTTTCAGCTGCTC 1020
QY 1021 CTTTCTCCGTGATCGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
Db 1021 CTTTCTCCGTGATCGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
QY 1081 TGACCCCTCCCGTCCGGTGGTTCGTATCAAGCTCTCAACCCCGTCCCGCGTGTGT 1140
Db 1081 TGACCCCTCCCGTCCGGTGGTTCGTATCAAGCTCTCAACCCCGTCCCGCGTGTGT 1140
QY 1141 CTTGCTGTGTGACGCTGCT 1200
Db 1141 CTTGCTGTGTGACGCTGCT 1200
QY 1201 GCTCGTGGGAGGTTCACCCCTTGGATCCAGGAAGAACCCCTCCACCCCTGCTCGTGGTG 1260
Db 1201 GCTCGTGGGAGGTTCACCCCTTGGATCCAGGAAGAACCCCTCCACCCCTGCTCGTGGTG 1260
QY 1261 GGCCAAAGGCTACAGGCTGCT 1320
Db 1261 GGCCAAAGGCTACAGGCTGCT 1320
QY 1321 TGGGCTGCTCTCCCGTACGCTGCGAAAGTGGAGGATCGAGGTAGGAGGAAACAGCAA 1380
Db 1321 TGGGCTGCTCTCCCGTACGCTGCGAAAGTGGAGGATCGAGGTAGGAGGAAACAGCAA 1380

QY 1381 CCGGGAGTCTCGAGCCTGGGGCTGCCCTACTCTACCCATTTCCCGACAGAGCTTTG 1440
Db 1381 CCGGGAGTCTCGAGCCTGGGGCTGCCCTACTCTACCCATTTCCCGACAGAGCTTTG 1440
QY 1441 CCTTGTCTGGCTGCCCGCTGCCTCTTTTGGGAACTGAGCTCAGAGGAGGTCTCAG 1500
Db 1441 CCTTGTCTGGCTGCCCGCTGCCTCTTTTGGGAACTGAGCTCAGAGGAGGTCTCAG 1500
QY 1501 AGAAGGAAACAAATGAGGGTGGCAGGGATAAAGTCAACCTCCCTCTCTACCTCCCA 1560
Db 1501 AGAAGGAAACAAATGAGGGTGGCAGGGATAAAGTCAACCTCCCTCTCTACCTCCCA 1560
QY 1561 TGCAGCATGAACACAAATTTCTCTCCACCTGGCTCCCAATTTAAAGATGTGACCAAGGC 1620
Db 1561 TGCAGCATGAACACAAATTTCTCTCCACCTGGCTCCCAATTTAAAGATGTGACCAAGGC 1620
QY 1621 CTGTGGGTACTCCAGGGCAAGGAGAGCCCTGGGGTCACTGACACTGTGAGGCCAACCAT 1680
Db 1621 CTGTGGGTACTCCAGGGCAAGGAGAGCCCTGGGGTCACTGACACTGTGAGGCCAACCAT 1680
QY 1681 GCACCTCAAAAGGGGAGCATTTGGAATGAAGGACTAGCTCTCTATGTATCAGGTTAAGA 1740
Db 1681 GCACCTCAAAAGGGGAGCATTTGGAATGAAGGACTAGCTCTCTATGTATCAGGTTAAGA 1740
QY 1741 GCAAGGGAGAGCTGGCCAGGGCAGCAGTTTGCACAGCAGAGGGGAATGTAGCAACAGCA 1800
Db 1741 GCAAGGGAGAGCTGGCCAGGGCAGCAGTTTGCACAGCAGAGGGGAATGTAGCAACAGCA 1800
QY 1801 GGGCTCTCTAGGCCCATCTTCCATTTCTAGGTAAGAGAGCATTTCTCTCAGACTCCCA 1860
Db 1801 GGGCTCTCTAGGCCCATCTTCCATTTCTAGGTAAGAGAGCATTTCTCTCAGACTCCCA 1860
QY 1861 GCGGAGGAGCTAGCCTAGCCTTACGAAACCAAGTTCTCTGGGACCCCAAGTTTATGG 1920
Db 1861 GCGGAGGAGCTAGCCTAGCCTTACGAAACCAAGTTCTCTGGGACCCCAAGTTTATGG 1920
QY 1921 GAGAAGGCAAGACTTTCATGGGAAGAGAGAGGAGGCCCTGGGTAGAACCTTGGTG 1980
Db 1921 GAGAAGGCAAGACTTTCATGGGAAGAGAGAGGAGGCCCTGGGTAGAACCTTGGTG 1980
QY 1981 CTGTTCTCTTGGCTTTTAAAGCAAGCGCTCATCTTGGCTCTACTCTCTCTGATAGGCTT 2040
Db 1981 CTGTTCTCTTGGCTTTTAAAGCAAGCGCTCATCTTGGCTCTACTCTCTCTGATAGGCTT 2040
QY 2041 GAGGTTTGCACACCACTGTGGCTACAGGTGAGGAGAGGAGTCTCTCTCCACAGAG 2100
Db 2041 GAGGTTTGCACACCACTGTGGCTACAGGTGAGGAGAGGAGTCTCTCTCCACAGAG 2100
QY 2101 TGTATGTTTCAGGAAGTTTCTTTAACCCCATATGGCCCAAGAGTGTCTGATAGGCCCC 2160
Db 2101 TGTATGTTTCAGGAAGTTTCTTTAACCCCATATGGCCCAAGAGTGTCTGATAGGCCCC 2160
QY 2161 TTTTAAAGCGGAACAAGTAAATTTACCAGTTCTACTGGGTTCTGCCCCACCGTCCCAAGG 2220
Db 2161 TTTTAAAGCGGAACAAGTAAATTTACCAGTTCTACTGGGTTCTGCCCCACCGTCCCAAGG 2220
QY 2221 TGGGCGAGGCTTAGGAAGGGTCTTTTAAAGCCACACATTAAGTGTGCTGCTGCTG 2280
Db 2221 TGGGCGAGGCTTAGGAAGGGTCTTTTAAAGCCACACATTAAGTGTGCTGCTGCTG 2280
QY 2281 CAGCCAAACAAAGACTGGGTGTGAGTATTCATCACTAAGAACCAAAATCCAGGGCA 2340
Db 2281 CAGCCAAACAAAGACTGGGTGTGAGTATTCATCACTAAGAACCAAAATCCAGGGCA 2340
QY 2341 CTCTATGTGAAGATAAGAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2400
Db 2341 CTCTATGTGAAGATAAGAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2400
QY 2401 ACCATCAAACTTTCT 2460
Db 2401 ACCATCAAACTTTCT 2460

Db 121 AGAAGGACTATATTTGTACTGTACCTAGGGTTCTGGAAGGAAACATGGAATCAGGA 180
QY 2862 TTCATAGACTGATAGGCCCTATCCACAAAGGCCATGACTGGAAAAGGTATGGAGCAG 2921
Db 181 TTCATAGACTGATAGGCCCTATCCACAAAGGCCATGACTGGAAAAGGTATGGAGCAG 240
QY 2922 AAGGAGAATTGGGATTTTAGGGTGCAGCTACGCTCACCTCAACCTTTTGGTGGCCTGGGG 2981
Db 241 AAGGAGAATTGGGATTTTAGGGTGCAGCTACGCTCACCTCAACCTTTTGGTGGCCTGGGG 300
QY 2982 CATGCTCTGAGGCCAGACTGTTAAGCAGGCTCTGCGCCCTGTTTACTGTCACACCT 3041
Db 301 YATGCTCTGAGGCCAGACTGTTAACCAGGCTCTGCGCCCTGTTTACTGTCACACCT 360
QY 3042 CTGACCTGCTGTTGAGACTCCATCCAGCCCCAGGCCACCTGCTCTGAGCCTC 3101
Db 361 CTGACCTGCTGTTGAGACTCCATCCAGCCCCAGGCCACCTGCTCTGAGCCTC 420
QY 3102 CACTATCTCCCTGTGACGGGTGAACCTTCTGTACTGTCTCGGTGCCATATATGAATTG 3161
Db 421 CACTATCTCCCTGTGACGGGTGAACCTTCTGTACTGTCTCGGTGCCATATATGAATTG 480
QY 3162 TGACGAGGTTCACTATATTTTAAACACAGATGTTTACAAAATAAGATTATTTCAAACA 3221
Db 481 TGACGAGGTTCACTATATTTTAAACACAGATGTTTACAAAATAAGATTATTTCAAACA 540
QY 3222 CC 3223
Db 541 CC 542

RESULT 6
US-09-817-198A-31
; Sequence 31, Application US/09817198A
; Patent No. US20020146758A1
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001188
; CURRENT APPLICATION NUMBER: US/09/817,198A
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-817-198A-31

Query Match 15.9%; Score 516.6; DB 10; Length 601;
Best Local Similarity 99.8%; Pred. No. 2.1e-99;
Matches 516; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2707 GGCAGTGGCTCCAAATCTGTGACACAGATTTTACGCTTTCCTGAAGATCAGGAGGTGC 2766
Db 1 GGCAGTGGCTCCAAATCTGTGACACAGATTTTACGCTTTCCTGAAGATCAGGAGGTGC 60
QY 2767 CATTCATGCTTCTCTCTAGCCCCCTCAGGAAGAAGGACTATATTTGTACTGTACC 2826
Db 61 CATTCATGCTTCTCTCTAGCCCCCTCAGGAAGAAGGACTATATTTGTACTGTACC 120
QY 2827 CTAGGGGTTCTGGAAGGAAACATGGAATCAGGATTTCTATAGACTGATAGGCCCTATCC 2886
Db 121 CTAGGGGTTCTGGAAGGAAACATGGAATCAGGATTTCTATAGACTGATAGGCCCTATCC 180
QY 2887 ACAAGGCCATGACTGGGAAAGATGTGGGACAGAGGAGATTTGGATTTTAGGGTGC 2946
Db 181 ACAAGGCCATGACTGGGAAAGATGTGGGACAGAGGAGATTTGGATTTTAGGGTGC 240
QY 2947 AGCTACGCTACCCCTAACTTTTGGTGGCCTGGGCGATGCTTTCAGGCCCCAGACTGTAA 3006

Db 241 AGCTACGCTACCCATAAATTTTGGTGGCCCTGGGGCATGTCTTGAGGCCCCAGACTGTAA 300
QY 3007 GCAGGCTCTGCTGGCCTGTTTACTCGTCACACACCTCTGCACCTCTGCTCTTGAGACTCCA 3066
Db 301 SCAGGCTCTGCTGGCCTGTTTACTCGTCACACACCTCTGCACCTCTGCTCTTGAGACTCCA 360
QY 3067 TCCAGGCCCCAGGACGCGCACCTGCTCTCGTGGCCCTCCACTATCTCCCTGTGAGGGTGAAC 3126
Db 361 TCCAGGCCCCAGGACGCGCACCTGCTCTCGTGGCCCTCCACTATCTCCCTGTGAGGGTGAAC 420
QY 3127 TTCGTGCTACTGCTCGGTGCCATATATGAATTGTGAGCAGGGTTCATCTATTTTAAAC 3186
Db 421 TTCGTGCTACTGCTCTCGGTGCCATATATGAATTGTGAGCAGGGTTCATCTATTTTAAAC 480
QY 3187 ACAGATGTTTACAAAATAAAGATTATTTCAAACACC 3223
Db 481 ACAGATGTTTACAAAATAAAGATTATTTCAAACACC 517

RESULT 7
US-09-920-300A-303
; Sequence 303, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 303
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 417..461
; OTHER INFORMATION: n = A,T,C or G
US-09-920-300A-303

Query Match 14.7%; Score 477.4; DB 10; Length 481;
Best Local Similarity 99.4%; Pred. No. 3.2e-91;
Matches 478; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2582 GTTGCTTCTCTTGAAGATGGGACTCCTTGGGTATCAAGACCTATGCCACATCACACTGGG 2641
Db 1 GTTGCTTCTCTTGAAGATGGGACTCCTTGGGTATCAAGACCTATGCCACATCACACTGGG 60
QY 2642 CTAGGGAAGTAGTGATGCCAGCCCTCAAGTCTGTCTTCAGCCAGGAGCTTGAGAAGTTA 2701
Db 61 CTAGGGAAGTAGTGATGCCAGCCCTCAAGTCTGTCTTCAGCCAGGAGCTTGAGAAGTTA 120
QY 2702 TATGGGCACTGGCTCCAAATCTGTGGACCAAGTATTTACGTTTCCCTGAAGTACGCG 2761
Db 121 TATGGGCACTGGCTCCAAATCTGTGGACCAAGTATTTACGTTTCCCTGAAGTACGCG 180
QY 2762 GGTGCCATTCATGCTCTTCTCTAGCCCCCTCAGGAAGAAGGACTATATTTTACT 2821
Db 181 GGTGCCATTCATGCTCTTCTCTAGCCCCCTCAGGAAGAAGGACTATATTTTACT 240
QY 2822 GTACCTAGGGTTCGGAAGGAAACATGGAATCAGGATTTCTATAGACTGATAGGCC 2881
Db 241 GTACCTAGGGTTCGGAAGGAAACATGGAATCAGGATTTCTATAGACTGATAGGCC 300
QY 2882 TATCCACAGGGCCATGACTGGGAAAGATGTGGGACAGAGGAAATTTGGATTTTAG 2941
Db 301 TATCCACAGGGCCATGACTGGGAAAGATGTGGGACAGAGGAAATTTGGATTTTAG 360

QY 2942 GGTGAGCTAGCTCACCTAAACTTTTGGTGGCTGGGGCATGCTTTGAGGCCAGACT 3001
|
Db 361 GGTGAGCTAGCTCACCTAAACTTTTGGTGGCTGGGGCATGCTTTGAGGCCANACT 420

QY 3002 GTTAAGCAGGCTCTGCTGGCCTGTTTACTCGTCACCACTCTGCACCTGCTGTCTTGAGA 3061
|
Db 421 GTTAACACGGCTCTGCTGGCCTGTTTACTCGTCACCACTCTGTCACCTGCTGTCTTGAGA 480

QY 3062 C 3062
|
Db 481 C 481

RESULT 8
US-10-033-528-303
; Sequence 303, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, JIangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033.528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 303
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 417, 461
; OTHER INFORMATION: n = A,T,C or G
US-10-033-528-303

Query Match 14.7%; Score 477.4; DB 12; Length 481;
Best Local Similarity 99.4%; Pred. No. 3.2e-91;
Matches 478; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2582 GTTGCTTCCTTGAAGTGGGACTCCTGGGTATCAAGACCTATGCCACATCACACTGGGG 2641
|
Db 1 GTTGCTTCCTTGAAGTGGGACTCCTGGGTATCAAGACCTATGCCACATCACACTGGGG 60

QY 2642 CTAGGGAAGTAGGTGATGCCAGCCCTCAAGTCTGCTTTACGCCAGGACTTGAGAAGTTA 2701
|
Db 61 CTAGGGAAGTAGGTGATGCCAGCCCTCAAGTCTGCTTTACGCCAGGACTTGAGAAGTTA 120

QY 2702 TATTGGGCAGTGGCTCCAACTGTGGACCACTATTTACAGTTTCCCTGAAGATCAGGCAG 2761
|
Db 121 TATTGGGCAGTGGCTCCAACTGTGGACCACTATTTACAGTTTCCCTGAAGATCAGGCAG 180

QY 2762 GGTGCATTATGTTCTTCTCTAGCCCTCAGGAAGAAGSACTATATTGTACT 2821
|
Db 181 GGTGCATTATGTTCTTCTCTAGCCCTCAGGAAGAAGSACTATATTGTACT 240

QY 2822 GTACCTTAGGGTTCCTGGAAGGAAACATGGAATCAGGATTCCTATAGACTGATAGGCC 2881
|
Db 241 GTACCTTAGGGTTCCTGGAAGGAAACATGGAATCAGGATTCCTATAGACTGATAGGCC 300

QY 2882 TATCCACAGGGCCATGACTGGGAAAGATATGGAGCAGAGGAATTTGGATTTTAG 2941
|
Db 301 TATCCACAGGGCCATGACTGGGAAAGATATGGAGCAGAGGAATTTGGATTTTAG 360

QY 2942 GGTGCAGCTACCTCACCTAACTTTTGGTGGCTGGGGCATGCTTTGAGGCCCAGACT 3001
|
Db 361 GGTGCAGCTACCTCACCTAACTTTTGGTGGCTGGGGCATGCTTTGAGGCCCANACT 420

QY 3002 GTTAAGCAGGCTCTGCTGGCCTGTTTACTCGTCACCACTCTGCACTGCTGTCTTGAGA 3061
|
|

Db 421 GTTAACACAGGCTCTGCTGGCCTGTTTACTCGTCACCACTCTGCACTGCTGTCTTGAGA 480

QY 3062 C 3062
|
Db 481 C 481

RESULT 9
US-09-964-824A-26/c
; Sequence 26, Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:
; APPLICANT: Horrigan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964.824A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236.033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236.032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236.028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
; LENGTH: 463
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-26

Query Match 13.5%; Score 438.8; DB 10; Length 463;
Best Local Similarity 99.3%; Pred. No. 3.9e-83;
Matches 451; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 2790 CCCCTCAGGAAAGGACTATATTGTACTGTACCTTAGGGTTCTGGAAGGAAAAAC 2849
|
Db 453 CCCCTCAGGAAAGGACTATATTGTACTGTACCTTAGGGTTCTGGAAGGAAAAAC 394

QY 2850 ATGGAATCAGGATTCCTATAGACTGATAGCCCTATCCACAAGGGCCATGACTGGGAAAG 2909
|
Db 393 ATGGAATCAGGATTCCTATAGACTGATAGCCCTATCCACAAGGGCCATGACTGGGAAAG 334

QY 2910 GTATGGAGCAGAGGAATTTGGATTTTAGGGTGCAGCTACGCTACCCCTAAACTTTT 2969
|
Db 333 GTATGGAGCAGAGGAATTTGGATTTTAGGGTGCAG-TACGCTCACCCCTAAACTTTT 275

QY 2970 GGTGGCCTGGGCACTGTCTTGAGGCCAGACTGTTAAGCAGGCTCTGCTGGCCTGTTTAC 3029
|
Db 274 GGTGGCCTGGGCACTGTCTTGAGGCCAGACTGTTAAGCAGGCTCTGCTGGCCTGTTTAC 215

QY 3030 TCGTCACCACTCTGCACCTGCTGTCTTGAGACTCCATCCAGCCCCCAGCAGCCACCCTG 3089
|
Db 214 TCGTCACCACTCTGCACCTGCTGTCTTGAGACTCCATCCAGCCCCCAGCAGCCACCCTG 155

QY 3090 CTCCTGAGCCTCCACTATCTCCCTGTGAGCGGTGAATTCGTGTACTGTGTCTCGGGTCC 3149
|
Db 154 CTCCTGAGCCTCCACTATCTCCCTGTGAGCGGTGAATTCGTGTACTGTGTCTCGGGTCC 95

QY 3150 ATATATGAATTCGTGAGCAGGTTTCATCTATTTTAAACACAGATGTTTACAAAATAAAGAT 3209
|
Db 94 ATATATGAATTCGTGAGCAGGTTTCATCTATTTTAAACACAGATGTTTACAAAATAAAGAT 35

QY 3210 TATTTTCAACACCAAAAAAATAAAAAA 3243
|
Db 34 TATTTTCAACACCAAAAAAATAAAAAA 1

RESULT 10
US-10-046-935-1930
; Sequence 1930, Application US/10046935
; Patent No. US20020156011A1

GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Harlocker, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1930
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-046-935-1930

Query Match 9.4%; Score 306.8; DB 9; Length 310;
Best Local Similarity 99.4%; Pred. No. 1.6e-55;
Matches 308; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2824 ACCTAGGGGTTCTGGAAGGAAACATGAATCAGATTCTATAGACTGATAGGCCCTA 2883
Db 1 ACCTAGGGGTTCTGGAAGGAAACATGAATCAGATTCTATAGACTGATAGGCCCTA 60
QY 2884 TCCACAGGGCCCATGCTGGGAAAGGTATGGGAGGAGGAGATTTGGGTTTAGGG 2943
Db 61 TCCACAGGGCCCATGCTGGGAAAGGTATGGGAGGAGGAGATTTGGGTTTAGGG 120
QY 2944 TGCAGCTACGCTCACCTAACTTTTGGTGGCCCTGGGGCATGCTTTGAGGCCGAGACTGT 3003
Db 121 TGCAGCTACGCTCACCTAACTTTTGGTGGCCCTGGGGCATGCTTTGAGGCCGAGACTGT 180
QY 3004 TAAGCAGGCTCTGCTGGCCCTGTTTACTGCTGCACACCTCTGCACCTCTGCTGTGAGACT 3063
Db 181 TAAGCAGGCTCTGCTGGCCCTGTTTACTGCTGCACACCTCTGCACCTCTGCTGTGAGACT 240
QY 3064 CCATCCAGCCCCAGGACGACCTGCTGCTGAGCCCTCCACTATCTCCCTGTGACGGGTG 3123
Db 241 CCATCCAGCCCCAGGACGACCTGCTGCTGAGCCCTCCACTATCTCCCTGTGACGGGTG 300

RESULT 11

US-09-794-257-9
; Sequence 9, Application US/09794257
; Patent No. US20020009804A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 32705, 23224, 27423, 32700, 32712, No. US20020009804A1el
; FILE REFERENCE: 35800/209285
; CURRENT APPLICATION NUMBER: US/09/794, 257
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/185,606
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 624
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-794-257-9

Query Match 6.7%; Score 217.8; DB 10; Length 624;
Best Local Similarity 65.6%; Pred. No. 9e-37;
Matches 318; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 45 ATGCGAAGCAGTACGATGTGCTTCCGGCTGCTGCTGATCGGGGACTCCGGGGTGGGC 104
Db 1 ATGCGAAGCAGTACGATGTGCTTCAAGCTCTGCTGCTGATCGGGGACTCCGGGGTGGGC 60
QY 105 AAGACCTGCCTGCTGCGGCTTACCGACACGAGTTCCACTCCTCGACATCTCCACC 164
Db 61 AAGACCTGCCTGCTGCGGCTTCTCAGAGGACGCTTCAACACCACTTCACTCTCCACC 120
QY 165 ATCGGTGTGACTTTAAGATGAAGACCATAGAGGTAGAGCGGATCAAAAGTTCGGATACAG 224
Db 121 ATCGGAATGTATTTAAATTTAGACGATAGAACTAGATGGAAGAAATTAAGCTTCAG 180
QY 225 ATCTGGACACTGCGAGGAGGAGATACAGACCATCACAAGACGAGTACTATCGGGG 284
Db 181 ATATGGACACAGCGGGTTCAGGAAAGATTCGGAACATACAGACGCTACTACAGAGA 240
QY 285 GCCAGGGGATATTTTGGTCTATGACATTTAGCAGGAGCGCTCTTACCAGCACATCATG 344
Db 241 GCATGGGCATTTATGCTGCTATGACATCACAATGAAATCTTTGCACATATTAAA 300
QY 345 AAGTGGGTGAGTACGCTGAGTACGACCAAGAGGCGTCCAGAAGATCCTTATTTGG 404
Db 301 AATGGATCAGAAACATTCAGAGCATGCTCTTCCGATGTCGAAAGATGATCCTGGGT 360
QY 405 AATAAGCTGATGAGGAGCAGAAACGGCAGGTGGGAGAGACCAAGGCGAGCGTGGG 464
Db 361 AACAATGTGATGATGAATGACAAAGACAAAGTGTCAAAAGAAAGAGGGGAGAGCTAGCA 420
QY 465 AAGGAGTATGGCATGACTTCTATGAACAAGTGCCTGCACCAACCTCAACATTAAGAG 524
Db 421 ATGACTATGGGATTAATTTCTTGGAGACAAGCGCAAAATCCAGTGCATAATAGAGAG 480
QY 525 TCATT 529
Db 481 GCATT 485

RESULT 12

US-09-794-257-7
; Sequence 7, Application US/09794257
; Patent No. US20020009804A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 32705, 23224, 27423, 32700, 32712, No. US20020009804A1el
; FILE REFERENCE: 35800/209285
; CURRENT APPLICATION NUMBER: US/09/794, 257
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/185,606
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (18)...(641)
US-09-794-257-7

Query Match 6.7%; Score 217.8; DB 10; Length 1161;
Best Local Similarity 65.6%; Pred. No. 1.1e-36;
Matches 318; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 45 ATGCGAAGCAGTACGATGTGCTTCCGGCTGCTGCTGATCGGGGACTCCGGGGTGGGC 104
Db 18 ATGCGAAGCAGTACGATTTCTTCAAGCTCTGCTGATCGGGGACTCCGGGGTGGGC 77
QY 105 AAGACCTGCCTGCTGCGGCTTACCGACCAAGAGTTCCACTCCTCGACATCTCCACC 164
Db 78 AAGACCTGCCTGCTTCCGGTTCAGAGGACGCTTCAACACCACTTCACTATCCACC 137

QY 165 ATCGGTGTTGACTTTAAGATGAAGACCATAGAGGTAGCGCATCAAAAGTCGGGATACAG 224
Db 138 ATCGGAATTGATTTTAAATTTAGAACGATAGAACTAGATGGAAGAAAATTAAGCTTCAG 197
QY 225 ATCTGGGACACTGCGAGGACGAGAGATACCAGACCATCAAAAGCAGTACTATCGGGG 284
Db 198 ATATGGGACACAGCGGGTCAGGAAGATTCGGAACAATCAGCAGCGTACTTACAGAGGA 257
QY 285 GCCAGGGGATATTTTGGTCTATGACATTTAGCAGCGAGCGTCTTACCAGCACATCATG 344
Db 258 GCCATGGCATATCTGCTGTATGACATCAAAATGAAATTCCTTTGACAAATTTAAA 317
QY 345 AAGTGGTCAGTGACGTGGATGAGTACGACCAAGCGGTCCAGAAGATCCTTATTGGG 404
Db 318 AATTGGATCAGAAACATTTGAAGAGCATGCCCTTCCGATGTCGAAAGAATGATCCTGGGT 377
QY 405 AATAAGGCTGATGAGGACAGAAAGCGGAGGTGGGAAGAGAGCAAGGGCAGCGTGGCG 464
Db 378 ACAAATGTGATGATGAATGACAAAAGACAAGTGTCAAAAGAAAGAGGAGAAAGCTAGCA 437
QY 465 AAGGATATGGCATGGACTTCTATGAAACAAGTGCCTGCACCAACCTCAACATTTAAAGAG 524
Db 438 ATTGACTATGGGATTAATTTCTTGGAGACAAGCGCAAAATCCAGTGCAAATGTAGAAGAG 497
QY 525 TCATT 529
Db 498 GCATT 502

RESULT 13
US-09-834-975-879
; Sequence 879, Application US/09834975
; Patent No. US20020110815A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Brown, Jeffrey
; APPLICANT: Bolt, Andrew
; APPLICANT: Van Huffel, Christophe
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
; TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF HUMAN CANCERS
; FILE REFERENCE: MRI-016B
; CURRENT APPLICATION NUMBER: US/09/834,975
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,538
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 1046
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 879
; LENGTH: 2497
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2497)
; OTHER INFORMATION: n = A,T,C or G
US-09-834-975-879

Query Match 6.7%; Score 217.8; DB 10; Length 2497;
Best Local Similarity 65.6%; Pred. No. 1.5e-36;
Matches 318; Conservative 0; Mismatches 167; Indels 0; Gaps 0;
QY 45 ATGGCGAAGCAGTACGATGTGCTGTCCGGCTGCTGCTGATCGGGGACTCCGGGTGGCG 104
Db 92 ATGGCGAAGCAGTACGATATCTCTCAAGCTCCTGTGATCGGCGACTCGGGGTAGGC 151
QY 105 AAGACTGCCTGCTGCGGCTTACCAGCAACAGAGTTCCTCCTCGACATCTCCACC 164
Db 152 AAGACTGCCTGCTGCTGCTTCTCAGAGGCGCTTCAACACCACCTTCACTCCACC 211
QY 165 ATCGGTGTTGACTTTAAGATGAAGACCATAGAGGTAGCGCATCAAAAGTCGGGATACAG 224
Db 212 ATCGGAATTGATTTTAAATTTAGAACGATAGAACTAGATGGAAGAAAATTAAGCTTCAG 271

QY 225 ATCTGGGACACTGCGAGGACGAGAGATACCAGACCATCAAAAGCAGTACTATCGGGG 284
Db 272 ATATGGGACACAGCGGTCAGGAAGATTCGGAACAATCAGCAGCGTACTTACAGAGA 331
QY 285 GCCAGGGGATATTTTGGTCTATGACATTTAGCAGCGAGCGTCTTACCAGCACATCATG 344
Db 332 GCCATGGCATATCTGCTGTATGACATCAAAATGAAAAATCCTTTGACAATATAAA 391
QY 345 AAGTGGTCAGTGACGTGGATGAGTACGACCAAGCGGTCCAGAAGATCCTTATTGGG 404
Db 392 AATTGGATCAGAAACATTTGAAGAGCATGCCCTTCCGATGTCGAAAGAATGATCCTGGGT 451
QY 405 AATAAGGCTGATGAGGACAGAAAGCGGAGGTGGGAAGAGAGCAAGGGCAGCGTGGCG 464
Db 452 RACAATGTGATGAATGACACAAAGACAAGTGTCAAAAGAAAGGGGAGAGCTAGCA 511
QY 465 AAGGATATGGCATGGACTTCTATGAAACAAGTGCCTGCACCAACCTCAACATTTAAAGAG 524
Db 512 ATTGACTATGGGATTAATTTCTTGGAGACAAGCGCAAAATCCAGTGCAAATGTAGAAGAG 571
QY 525 TCATT 529
Db 572 GCATT 576

RESULT 14
US-09-834-975-885
; Sequence 885, Application US/09834975
; Patent No. US20020110815A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Brown, Jeffrey
; APPLICANT: Bolt, Andrew
; APPLICANT: Van Huffel, Christophe
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
; TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF HUMAN CANCERS
; FILE REFERENCE: MRI-016B
; CURRENT APPLICATION NUMBER: US/09/834,975
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,538
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 1046
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 885
; LENGTH: 2497
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2497)
; OTHER INFORMATION: n = A,T,C or G
US-09-834-975-885
Query Match 6.7%; Score 217.8; DB 10; Length 2497;
Best Local Similarity 65.6%; Pred. No. 1.5e-36;
Matches 318; Conservative 0; Mismatches 167; Indels 0; Gaps 0;
QY 45 ATGGCGAAGCAGTACGATGTGCTGTCCGGCTGCTGCTGATCGGGGACTCCGGGTGGCG 104
Db 92 ATGGCGAAGCAGTACGATATCTCTCAAGCTCCTGTGATCGGCGACTCGGGGTAGGC 151
QY 105 AAGACTGCCTGCTGCGGCTTACCAGCAACAGAGTTCCTCCTCGACATCTCCACC 164
Db 152 AAGACTGCCTGCTGCTGCTTCTCAGAGGACGCGCTTCAACACCACCTTCACTCCACC 211
QY 165 ATCGGTGTTGACTTTAAGATGAAGACCATAGAGGTAGCGCATCAAAAGTCGGGATACAG 224
Db 212 ATCGGAATTGATTTTAAATTTAGAACGATAGAACTAGATGGAAGAAAATTAAGCTTCAG 271
QY 225 ATCTGGGACACTGCGAGGACGAGAGATACCAGACCATCAAAAGCAGTACTATCGGGG 284
Db 272 ATATGGGACACAGCGGTCAGGAAGATTCGGAACAATCAGCAGCGTACTTACAGAGA 331
QY 285 GCCAGGGGATATTTTGGTCTATGACATTTAGCAGCGAGCGTCTTACCAGCACATCATG 344
Db 332 GCCATGGCATATCTGCTGTATGACATCAAAATGAAAAATCCTTTGACAATATAAA 391
QY 345 AAGTGGTCAGTGACGTGGATGAGTACGACCAAGCGGTCCAGAAGATCCTTATTGGG 404
Db 392 AATTGGATCAGAAACATTTGAAGAGCATGCCCTTCCGATGTCGAAAGAATGATCCTGGGT 451
QY 405 AATAAGGCTGATGAGGACAGAAAGCGGAGGTGGGAAGAGAGCAAGGGCAGCGTGGCG 464
Db 452 RACAATGTGATGAATGACACAAAGACAAGTGTCAAAAGAAAGGGGAGAGCTAGCA 511
QY 465 AAGGATATGGCATGGACTTCTATGAAACAAGTGCCTGCACCAACCTCAACATTTAAAGAG 524
Db 512 ATTGACTATGGGATTAATTTCTTGGAGACAAGCGCAAAATCCAGTGCAAATGTAGAAGAG 571
QY 525 TCATT 529
Db 572 GCATT 576

Db 272 ATATGGGACACACGCGGTTCAGGAAGATTCCGAACAATCACACAGCGCTACTACAGAGGA 331
QY 285 GCCCAGGGGATATTTTGTCTATGACATTAGCAGCGCGCTCTTACCAGCACATCATG 344
Db 332 GCCATGGGCAATTATGCTGTCTATGACATCAACAATGAAAAATCCTTGCACAATATTAA 391
QY 345 AAGTGGGTCAAGTACGTGAGTACGCACCAAGAGGCGTCCAGAAGATCCTTATTGGG 404
Db 392 AATTGGATCAGAAACATTGAAGAGCATGCCTCTCCGATGTCGAAAGATGATCCTGGGT 451
QY 405 AATAAGGCTGATGAGGAGCAGAAACGGCAGGTGGGAAGAGCAAGGGCAGCAGCTGGCG 464
Db 452 AACAAATGTGATGAATGACAAAAAGACAAGTGTCAAAAAGAAAGAGGGGAGAGCTAGCA 511
QY 465 AAGGAGTATGGCATGGACTTCTATGAACAAGTGCCTGCACCAACCTCAACATTAAAGAG 524
Db 512 ATTGACTATGGGATTAAATTTCTTGAGAGACAGCGCAAAATCCAGTCCAAATGTAGAGAG 571
QY 525 TCATT 529
Db 572 GCATT 576

RESULT 15
US-09-834-975-894
; Sequence 894, Application US/09834975
; Patent No. US20020110815A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Brown, Jeffrey
; APPLICANT: Bolt, Andrew
; APPLICANT: Van Huffel, Christophe
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
; TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF HUMAN CANCERS
; FILE REFERENCE: MRI-016B
; CURRENT APPLICATION NUMBER: US/09/834,975
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,538
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 1046
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 894
; LENGTH: 2497
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2497)
; OTHER INFORMATION: n = A,T,C or G
US-09-834-975-894

Query Match 6.7%; Score 217.8; DB 10; Length 2497;
Best Local Similarity 65.6%; Pred. No. 1.5e-36;
Matches 318; Conservative 0; Mismatches 167; Indels 0; Gaps 0;
QY 45 ATGGCGAAGCAGTACGATGTGCTGTTCCGGCTGCTGTCGCGGAGCTCCGGGGTGGGC 104
Db 92 ATGGCGAAGCAGTACGATTTATCTTCAAGCTCTCTGCTGATCGCGAGCTCGGGGTAGGC 151
QY 105 AAGACCTGCTGTGTCGCCCTTCCAGGACACAGAGTTCACCTCCCTCCACATCTCCACC 164
Db 152 AAGACCTGCTGCTGTGTTCCGCTTCTCAGAGGAGCGCTTCAACACCCACCTTTCATCTCAOC 211
QY 165 ATCGGTGTTGACTTTAAGATGAGACCATAGAGGTAGCGGCATCAAAAGTCCGATACAG 224
Db 212 ATCGGAATTTGATTTTAAATTTAGACGTAGAACTAGATGGAAGAAATTAAGCTTCAG 271
QY 225 ATCTGGACACTCGAGGCGAGAGATACCAAGCAACATCAAAAGCAGTACTATCGGGCGG 284
Db 272 ATATGGACACACGCGGTTCAGGAAGATTCCGAACAATCACGACAGCGTACTACAGAGGA 331
QY 285 GCCCAGGGGATATTTTGTCTATGACATTAGCAGCGCGCTCTTACCAGCACATCATG 344

Db 332 GCCATGGGCAATTATGCTGTCTATGACATCAACAATGAAAAATCCTTGCACAATATTAA 391
QY 345 AAGTGGGTCAAGTACGTGAGTACGCACCAAGAGGCGTCCAGAAGATCCTTATTGGG 404
Db 392 AATTGGATCAGAAACATTGAAGAGCATGCCTCTCCGATGTCGAAAGATGATCCTGGGT 451
QY 405 AATAAGGCTGATGAGGAGCAGAAACGGCAGGTGGGAAGAGCAAGGGCAGCAGCTGGCG 464
Db 452 AACAAATGTGATGAATGACAAAAAGACAAGTGTCAAAAAGAAAGAGGGGAGAGCTAGCA 511
QY 465 AAGGAGTATGGCATGGACTTCTATGAACAAGTGCCTGCACCAACCTCAACATTAAAGAG 524
Db 512 ATTGACTATGGGATTAAATTTCTTGAGAGACAGCGCAAAATCCAGTCCAAATGTAGAGAG 571
QY 525 TCATT 529
Db 572 GCATT 576

Search completed: November 17, 2002, 20:38:39
Job time : 194 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2002, 16:43:38 ; Search time 434 Seconds
(without alignments)
16900.365 Million cell updates/sec

Title: US-09-817-198A-1
Perfect score: 3257
Sequence: 1 tgcccgctgccgcgcgcag.....aaaaaaaaaaaaaaaa 3257

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	2696.6	82.8	3240	22	AAK82285
2	1568.2	48.1	1666	21	AACT5813
3	1192.8	36.6	7924	24	ABK40069
4	1192.8	36.6	7924	24	ABL34138
C 5	1160.8	35.6	7924	24	ABK40070
C 6	1160.8	35.6	7924	24	ABL34139
7	1160.6	35.6	2021	22	AACT5053
8	866.2	26.6	895	20	AAZ24400
9	534.2	16.4	566	22	AAS27472

10	477.4	14.7	481	24	ABK44752
11	438.8	13.5	463	24	ABL66992
12	313.4	9.6	320	16	AAT25457
13	307	9.4	313	24	ABL38341
14	229	7.0	1986	22	ABA09160
15	228.4	7.0	2411	24	ABQ54410
16	217.8	6.7	1161	22	AAH75182
17	217.8	6.7	2497	22	AAS60878
18	217.8	6.7	2497	22	AAS60884
19	217.8	6.7	2497	22	AAS60893
20	217.8	6.7	2497	22	AAS60895
21	217.8	6.7	2497	23	ABV25781
22	217.8	6.7	2497	23	ABV30037
23	217.8	6.7	3077	22	AAH13912
24	214.2	6.6	866	22	AAH04301
25	202.6	6.2	412	21	AAC01700
26	200.2	6.1	716	21	AAA40104
27	200.2	6.1	861	21	AAA40108
28	198.4	6.1	956	21	AAA96887
29	198.4	6.1	1537	21	AAF16196
30	198.4	6.1	3533	22	AAH17889
31	196.8	6.0	674	23	AAS71453
32	196.6	6.0	2247	23	ABL29661
33	195.2	6.0	888	21	AAA40109
34	192	5.9	1274	21	AAF18072
35	188.6	5.8	893	23	AAS87552
36	185.4	5.7	1540	23	ABL27707
37	174.2	5.3	730	21	AAS26696
38	173.4	5.3	1023	21	AAC51491
39	173.2	5.3	742	24	ABK83539
40	171.8	5.3	1025	21	AAC34347
41	166.2	5.1	1203	21	AAC43685
42	165.4	5.1	1112	21	AAC46645
43	160.8	4.9	1098	21	AAC48038
44	159.2	4.9	1518	24	ABK71569
45	156.8	4.8	893	23	AAS76826

ALIGNMENTS

RESULT 1

AAK82285/c	
ID	AAK82285 standard; DNA; 3240 BP.
XX	
AC	AAK82285;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37097.
DE	
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW	cytostatic; gene therapy; vaccine; metastasis; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200157182-A2.
XX	
PD	09-AUG-2001.
XX	
PF	
XX	17-JAN-2001; 2001WO-US01354.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-02116647.
PR 07-JUL-2000; 2000US-02116880.
PR 11-JUL-2000; 2000US-02117487.
PR 11-JUL-2000; 2000US-02117496.
PR 14-JUL-2000; 2000US-02118290.
PR 26-JUL-2000; 2000US-02209063.
PR 26-JUL-2000; 2000US-02209664.
PR 14-AUG-2000; 2000US-02245118.
PR 14-AUG-2000; 2000US-02245119.
PR 14-AUG-2000; 2000US-02252113.
PR 14-AUG-2000; 2000US-02252114.
PR 14-AUG-2000; 2000US-02252666.
PR 14-AUG-2000; 2000US-02252667.
PR 14-AUG-2000; 2000US-02252668.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0244676.
PR 08-NOV-2000; 2000US-0244677.
PR 08-NOV-2000; 2000US-0244678.
PR 08-NOV-2000; 2000US-024523.
PR 08-NOV-2000; 2000US-024524.
PR 08-NOV-2000; 2000US-024525.
PR 08-NOV-2000; 2000US-024526.
PR 08-NOV-2000; 2000US-024527.
PR 08-NOV-2000; 2000US-024528.
PR 08-NOV-2000; 2000US-024532.
PR 08-NOV-2000; 2000US-024532.
PR 08-NOV-2000; 2000US-0245609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX Disclosure; SEQ ID NO 37097; 3071pp + Sequence Listing; English.
PS
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to

CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 3240 BP; 713 A; 882 C; 887 G; 758 T; 0 other;

Query Match 82.8%; Score 2596.6; DB 22; Length 3240;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 2717; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY	473	TGCGATGGACTTCTATGAACAAGTCCCTGACCAACCTCAACATTAAAGAGTCATTAC	532
DB	3100	TGCCAGGGAGAGGAGGACACTGGACTAACCTGTGCCCTTTGGTTTCCAGTCATTAC	3041
QY	533	GGCTCTGACAGAGCTGGTCTCAGGCCCATAGGAAGAGCTGGAAGGCTCCGGATGG	592
DB	3040	GGCTCTGACAGAGCTGGTCTCAGGCCCATAGGAAGAGCTGGAAGGCTCCGGATGG	2981
QY	593	TGCCAGCAATAGTTGGCACTGGCAGAGCTGGAGAGGAGGCAACACCGAGGGGCC	652
DB	2980	TGCCAGCAATAGTTGGCACTGGCAGAGCTGGAGAGGAGGCAACACCGAGGGGCC	2921
QY	653	ACGGAACCTCTCGAAACCTGCTGTGCTGAGTCTGTGGGGCACCACACACACACC	712
DB	2920	ACGGAACCTCTCGAAACCTGCTGTGCTGAGTCTGTGGGGCACCACACACACACC	2861
QY	713	CCTCTTCCCTCAGGAGGCCGCTGGGCAGACAGGGGAGCGGGGCTTGGCCCTGCTGCTG	772
DB	2860	CCTCTTCCCTCAGGAGGCCGCTGGGCAGACAGGGGAGCGGGGCTTGGCCCTGCTGCTG	2801
QY	773	CCTCTGCTGATGACCTTATTGATATCAGTAGCCACTACTCCCTGCTGGCCCTGA	832
DB	2800	CCTCTGCTGATGACCTTATTGATATCAGTAGCCACTACTCCCTGCTGGCCCTGA	2741
QY	833	GAGCGGCTCTGCTGATCTCAAGCAGCCCTGTCCCGACCGCGTCCACCTTGGAGTGT	892
DB	2740	GAGCGGCTCTGCTGATCTCAAGCAGCCCTGTCCCGACCGCGTCCACCTTGGAGTGT	2681
QY	893	CTTCTTACGCTGTTTCCCGACCGACAGCGCTGTACGACCCCGCAGATGTCGCGCAAGC	952
DB	2680	CTTCTTACGCTGTTTCCCGACCGACAGCGCTGTACGACCCCGCAGATGTCGCGCAAGC	2621
QY	953	ACTGTCTACCAATCCCGCACCCACAGACACAGCCAGGGCTGGAGTCCAGGCCACTTTC	1012
DB	2620	ACTGTCTACCAATCCCGCACCCACAGACACAGCCAGGGCTGGAGTCCAGGCCACTTTC	2561
QY	1013	AGCTGCTCCTTCTCCGTGCATCGTCTCTCTCTGCTTTTCTCTCTCTCTCTCTCTCT	1072
DB	2560	AGCTGCTCCTTCTCCGTGCATCGTCTCTCTCTCTGCTTTTCTCTCTCTCTCTCTCT	2501
QY	1073	TCTTTCTCTGACCCCTCCCTCCGCTGGGTTTCGTATCAAAAGCTCCTCAAAACCCGTC	1132
DB	2500	TCTTTCTCTGACCCCTCCCTCCGCTGGGTTTCGTATCAAAAGCTCCTCAAAACCCGTC	2441
QY	1133	CGGTGTCTCTGCTGTGACGCTCGCTCTTTCCTCTCTCTCTCTCTCTCTCTCTCTCT	1192
DB	2440	CGGTGTCTCTGCTGTGACGCTCGCTCTTTCCTCTCTCTCTCTCTCTCTCTCTCTCT	2381
QY	1193	GGACCCAGGCTCGTGGGAGGTTTCCACCTTGGATCCAGGAAGACCTCCACCTGGCT	1252
DB	2380	GGACCCAGGCTCGTGGGAGGTTTCCACCTTGGATCCAGGAAGACCTCCACCTGGCT	2321
QY	1253	CGTGGTGGGCAAAAGGCTTACAGGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1312
DB	2320	CGTGGTGGGCAAAAGGCTTACAGGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2261
QY	1313	ATGTGCCATGGGCTTGCCTCCCGAGTGACCTGCGAAAGTGGAGCATCGAGGTAGGAGGA	1372

DB	2260	ATGTGCCATGGGCTGCTCCCTCCCGAGTGACCTGCGAAAGTGGAGCATCGAGGTAGGAGGA	2201
QY	1373	ACAGCAACCCGGGAGTCTCTCAGGCTGGGCTGCCCTACCTTACCTTACCTTACCTTACCT	1432
DB	2200	AACGGCAACGAGGAGTCTCTCAGGCTGGGCTGCCCTACCTTACCTTACCTTACCTTACCT	2141
QY	1433	GAGCTTTGGCTTGTCTGGCTGCCGCTGCTCTTTGGGGAAGTCTGAGCTCAGAGCAGG	1492
DB	2140	GAGCTTTGGCTTGTCTGGCTGCCGCTGCTCTTTGGGGAAGTCTGAGCTCAGAGCAGG	2081
QY	1493	TGCTTCAGAGAAGAAACAAATGAGGGTGGCAGGATATAAAGTCACTCCATTTCTCT	1552
DB	2080	TGCTTCAGAGAAGAAACAAATGAGGGTGGCAGGATATAAAGTCACTCCATTTCTCT	2021
QY	1553	ACCTCCCATGACGATGAACAAATTTCTCTCCACCTGCTCCCAATTTAAAGATGTGG	1612
DB	2020	ACCTCCCATGACGATGAACAAATTTCTCTCCACCTGCTCCCAATTTAAAGATGTGG	1961
QY	1613	ACCAAGGCTGTGGGTACTTCCAGGGGCAAGAGAGCCCTGGGCTCAGTGACACTGTCAAG	1672
DB	1960	ACCAAGGCTGTGGGTACTTCCAGGGGCAAGAGAGCCCTGGGCTCAGTGACACTGTCAAG	1901
QY	1673	CCAACCATGCATCTCCACAAAGGGGAGCATTTGGAATGAAGGACTAGCTTCTATGTATCA	1732
DB	1900	CCAACCATGCATCTCCACAAAGGGGAGCATTTGGAATGAAGGACTAGCTTCTATGTATCA	1841
QY	1733	GGTTAAGAGCAGGAGAGCTGCCAGGAGCAGCAGTTTGCACAGCAGAGGGAATGTAG	1792
DB	1840	GGTTAAGAGCAGGAGAGCTGCCAGGAGCAGCAGTTTGCACAGCAGAGGGAATGTAG	1781
QY	1793	CAACAGAGGCTCTCTAGGCCCCATCTTCCATTTCTTAAAGTAAAGAGCATTCTCTCA	1852
DB	1780	CAACAGAGGCTCTCTAGGCCCCATCTTCCATTTCTTAAAGTAAAGAGCATTCTCTCA	1721
QY	1853	GACTCCCGGCGGAGCTAGCGCTTACAGCAACCAAGGTTCTCTGGGACCCAAA	1912
DB	1720	GACTCCCGGCGGAGCTAGCGCTTACAGCAACCAAGGTTCTCTGGGACCCAAA	1661
QY	1913	GTTTATGGAGAGGCAAGACTTTCATGGGAGAGAGAGGAGCCCTGGGTAGAAC	1972
DB	1660	GTTTATGGAGAGGCAAGACTTTCATGGGAGAGAGAGGAGCCCTGGGTAGAAC	1601
QY	1973	GGTTGTGCTGTTCTTTGGCTTTAAGCAAAAGCGCTCATCTTCCCTCTACCTCTCG	2032
DB	1600	GGTTGTGCTGTTCTTTGGCTTTAAGCAAAAGCGCTCATCTTCCCTCTACCTCTCG	1541
QY	2033	ATAGGCTTGGGTTTGGCAACACACTGTGCTACAGTGGAGGAGGAGGACTCTCTTC	2092
DB	1540	ATAGGCTTGGGTTTGGCAACACACTGTGCTACAGTGGAGGAGGAGGACTCTCTTC	1481
QY	2093	CTCCAGAGTGTATGTTAGGAGTTCCTTAAACCCATATGCCCAAGAGTAGTCGTA	2152
DB	1480	CTCCAGAGTGTATGTTAGGAGTTCCTTAAACCCATATGCCCAAGAGTAGTCGTA	1421
QY	2153	GGAGGCTTTAAAGACGGAACAAGTAATTTACAGTTCCTGAGTTCCTGCCACCG	2212
DB	1420	GGAGGCTTTAAAGACGGAACAAGTAATTTACAGTTCCTGAGTTCCTGCCACCG	1361
QY	2213	TCCCAAGTGGCGAGGCTTAGGAAGGTCATTTTAAAGCCACATTAAGTGCACATG	2272
DB	1360	TCCCAAGTGGCGAGGCTTAGGAAGGTCATTTTAAAGCCACATTAAGTGCACATG	1301
QY	2273	CGTGGCTGCAGCCAAACAAAGAACTGGGTGTGATTTATCACTAAAGAACCAAAAT	2332
DB	1300	CGTGGCTGCAGCCAAACAAAGAACTGGGTGTGATTTATCACTAAAGAACCAAAAT	1241
QY	2333	CCAGGCACTCATATGTGAAGGATAAGAACCTCACTTCTTACTCTCCAAAAGAGTG	2392
DB	1240	CCAGGCACTCATATGTGAAGGATAAGAACCTCACTTCTTACTCTCCAAAAGAGTG	1181
QY	2393	GGGAAAGAACCATCAAAACCTTTCTCTCTGACTTACCAACAGGAAACAGCAGGAGG	2452

Db 1180 GGAAGAAGAACCTCAACCTTCTCTCTGACTTACCAACACAGGAAACACGACGAGGAGG 1121
QY 2453 GTGGCTCAGACTTAGGACAGGGTATAGCTTAGATGGTGGAAAGCAAGAGGACGACGA 2512
Db 1120 GTGGCTCAGACTTAGGACAGGGTATAGCTTAGATGGTGGAAAGCAAGAGGACGACGA 1061
QY 2513 AGTTGTAATCACTGGCTAATGAGAAAGGAGACACTTAACCTAGATGAAGCTGTGAC 2572
Db 1060 AGTTGTAATCACTGGCTAATGAGAAAGGAGACACTTAACCTAGATGAAGCTGTGAC 1001
QY 2573 TAGGCTGGAGTTCCTTCCCTTGAAGATGGGACTCCTTGGGTATCAAGACCTATGCCACATC 2632
Db 1000 TAGGCTGGAGTTCCTTCCCTTGAAGATGGGACTCCTTGGGTATCAAGACCTATGCCACATC 941
QY 2633 ACACCTGGGCTAGGGAAGTAGGTGATGCCAGCCCTCAAGTCTGTCTTCAGCCAGGGACTT 2692
Db 940 ACACCTGGGCTAGGGAAGTAGGTGATGCCAGCCCTCAAGTCTGTCTTCAGCCAGGGACTT 881
QY 2693 GAGAAGTTATATGGGCAGTGGCTCAANTCTGTGGACCAAGTATTTTCAGCTTTCCTTGAAG 2752
Db 880 GAGAAGTTATATGGGCAGTGGCTCAANTCTGTGGACCAAGTATTTTCAGCTTTCCTTGAAG 821
QY 2753 ATCAGCAGGGTGCATTCATCTCTTCTCTCTAGCCCTCAGGAAAGAGGACTAT 2812
Db 820 ATCAGCAGGGTGCATTCATCTCTTCTCTAGCCCTCAGGAAAGAGGACTAT 761
QY 2813 ATTTGTAATCACTAGGCTTCTGGAAGGAAACATGGAATCAGGATTTCTATAGACT 2872
Db 760 ATTTGTAATCACTAGGCTTCTGGAAGGAAACATGGAATCAGGATTTCTATAGACT 701
QY 2873 GATAGCCCTATCCACAGGGCCATGACTGGGAAAGGATGAGGACGAGGAATG 2932
Db 700 GATAGCCCTATCCACAGGGCCATGACTGGGAAAGGATGAGGACGAGGAATG 641
QY 2933 GATATTTAGGTCAGCTACGCTACCCCTAAACTTTTGTGGCTGGGCACTGCTTGAG 2992
Db 640 GATATTTAGGTCAGCTACGCTACCCCTAAACTTTTGTGGCTGGGCACTGCTTGAG 581
QY 2993 GCCAGACTGTTAAGCAGGCTGCTGGGCTGTTTACTCGTACACCTCTGCACCTGCT 3052
Db 580 GCCAGACTGTTAAGCAGGCTGCTGGGCTGTTTACTCGTACACCTCTGCACCTGCT 521
QY 3053 GTCTTGAGACTCATCCAGCCAGGACGACCCACCTGCTCGTACGCTTCGATCTCC 3112
Db 520 GTCTTGAGACTCATCCAGCCAGGACGACCCACCTGCTCGTACGCTTCGATCTCC 461
QY 3113 TGTGAGGCTGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3172
Db 460 TGTGAGGCTGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 401
QY 3173 CATCTATTTTAAACAGATGTTTACAAAATAAGATTTTCAAAACCACC 3223
Db 400 CATCTATTTTAAACAGATGTTTACAAAATAAGATTTTCAAAACCACC 350

RESULT 2
AAC75813
ID AAC75813 standard; cDNA; 1666 BP.
XX
AC AAC75813;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1368 polynucleotide sequence SEQ ID NO:2735.
XX
DE Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
KW vulnerable; antiparasitic; antiparasitic; antiparasitic; hepatotropic;
KW anticonvulsant; osteopathic; antihypertensive; immunosuppressive; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antihypertensive;
KW antiviral; antibacterial; antifungal; antihypertensive; antihypertensive;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX Homo sapiens.
OS
PN WO200058473-A2.
XX
XX 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB41604.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 1978-1979; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiparasitic; antiparasitic; antiparasitic; nontropic; neuroprotective;
CC osteopathic; anticonvulsant; antihypertensive; immunosuppressive;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; antihypertensive;
CC antiinflammatory; antibacterial; antiviral; immunosuppressive;
CC antihypertensive; antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus, lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 1666 BP; 333 A; 529 C; 461 G; 341 T; 2 other;

Query Match 48.1%; Score 1568.2; DB 21; Length 1666;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 1598; Conservative 0; Mismatches 3; Indels 18; Gaps 1;
QY 2 GCCCGCTGCCCGCCGAGTTCGCCCGCCCGCTGGCCCCCAGTATGCGGAGCAGTACGA 61
Db 48 GCCCGCTGCCCGCCGAGTTCGCCCGCCCGCTGGCCCCCAGTATGCGGAGCAGTACGA 107
QY 62 TGTGCTGTTCCGCTGCTGCTGATCGGGGACTCCGGGTTGGGAGACCTGCTGCTGTG 121
Db 108 TGTGCTGTTCCGCTGCTGCTGATCGGGGACTCCGGGTTGGGAGACCTGCTGCTGTG 167
QY 122 CCCTCTTCCACGACAGGAGTTCACCTCTCTCGCACATCTCCACCATCGGTGTTGACTTTAA 181
Db 168 CCCTCTTCCACGACAGGAGTTCACCTCTCTCGCACATCTCCACCATCGGTGTTGACTTTAA 227

Qy 182 GATGAAGACCATAGAGGTAGACGGCATCAAAAGTCGGGATACAGATCTGGGACATCGAGG 241
Db 228 GATGAAGACCATAGAGGTAGACGGCATCAAAAGTCGGGATACAGATCTGGGACATCGAGG 287
Qy 242 GCAGGAGATACAGACCATCAAAAGTCAGTACTATCGGGGCCCGAGGGGATATTTT 301
Db 288 GCAGGAGATACAGACCATCAAAAGTCAGTACTATCGGGGCCCGAGGGGATATTTT 347
Qy 302 GGTCTATGACATTAGCAGCGAGCGCTCTTACCAGCACATCATGAAGTGGGTGAGTGCAGT 361
Db 348 GGTCTATGACATTAGCAGCGAGCGCTCTTACCAGCACATCATGAAGTGGGTGAGTGCAGT 407
Qy 362 GGATGAGTACGCACAGAGCGGTCCAGAGATCCTTATTTGGGAATTAAGGCTGATGAGGA 421
Db 408 GGATGAGTACGCACAGAGCGGTCCAGAGATCCTTATTTGGGAATTAAGGCTGATGAGGA 467
Qy 422 GCAGAAAGCGAGGTGGGAAGAGAGCAAGGCGAG-----CAGCTGGC 463
Db 468 GCAGAAAGCGAGGTGGGAAGAGAGCAAGGCGAGAAATGTCTCTTCTTCAGCTGGC 527
Qy 464 GAAGGAGTATGGCATGGACTTCTATGAACAAGTGCCTGCACCAACCTCAACATTAAGA 523
Db 528 GAAGGAGTATGGCATGGACTTCTATGAACAAGTGCCTGCACCAACCTCAACATTAAGA 587
Qy 524 GTCATTACGCGTCTGACAGAGCTGGTCTGACAGGCCCATAGGAAGAGCTGGAAGCGCT 583
Db 588 GTCATTACGCGTCTGACAGAGCTGGTCTGACAGGCCCATAGGAAGAGCTGGAAGCGCT 647
Qy 584 CCGGATCGGTGCCAGCATGAGTTGGCACATGGCAGAGCTGGAGAGGAGGCGCAAC 643
Db 648 CCGGATCGGTGCCAGCATGAGTTGGCACATGGCAGAGCTGGAGAGGAGGCGCAAC 707
Qy 644 CGAGGGCCACGCAACTCTTCGAAAACCTGCTGTGCTGAGTCTGTGTGGGCGACCCCA 703
Db 708 CGAGGGCCACGCAACTCTTCGAAAACCTGCTGTGCTGAGTCTGTGTGGGCGACCCCA 767
Qy 704 CACGACACCCCTCTTCCCTCAGGAGGCCCGTGGCAGACAGGGGAGCGGGGCTTTGCC 763
Db 768 CACGACACCCCTCTTCCCTCAGGAGGCCCGTGGCAGACAGGGGAGCGGGGCTTTGCC 827
Qy 764 TGCTGCTGCTCTCGTGTGATGACCCCTATTGAGTATCAGTACGACTACTCTCCCTGCC 823
Db 828 TGCTGCTGCTCTCGTGTGATGACCCCTATTGAGTATCAGTACGACTACTCTCCCTGCC 887
Qy 824 TGGCCCTGAGAGCGCTGCTGTCTCATCTCAAGCAGCCCTGTCCCGAGCCCGTCCACCC 883
Db 888 TGGCCCTGAGAGCGCTGCTGTCTCATCTCAAGCAGCCCTGTCCCGAGCCCGTCCACCC 947
Qy 884 TGGAGTGGTCTTTCTCAGCCCTGTTTCCCGCAGCCACAGGCTGTCTACGACCCCGCATGT 943
Db 948 TGGAGTGGTCTTTCTCAGCCCTGTTTCCCGCAGCCACAGGCTGTCTACGACCCCGCATGT 1007
Qy 944 GCGGAGAGCTGTCTCACCATCCCGCAGCCACAGCAGCAGCGGCTGGAGTCCAG 1003
Db 1008 GCGGAGAGCTGTCTCACCATCCCGCAGCCACAGCAGCAGCGGCTGGAGTCCAG 1067
Qy 1004 GCCACTTTCAGCTGCTCTTCTCCGTCGATCCTCTCTCTCTCTCTCTCTCTCTCTCTC 1063
Db 1068 GCCACTTTCAGCTGCTCTTCTCCGTCGATCCTCTCTCTCTCTCTCTCTCTCTCTCTC 1127
Qy 1064 CCCCACTTCTCTCTCTGACCCCTCCCGTCCGCTGGGTTTCGTATCAAAAGCTCCTCAA 1123
Db 1128 CCCCACTTCTCTCTCTGACCCCTCCCGTCCGCTGGGTTTCGTATCAAAAGCTCCTCAA 1187
Qy 1124 CCCCGTCCCGCTGTGCTGCTGTGAGCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1183
Db 1188 CCCCGTCCCGCTGTGCTGCTGTGAGCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1247
Qy 1184 CAAGGAGTACACCGCTCGTGGGAGGTTCCACCTTGTGATCCAGGAGAACCTCC 1243
Db 1248 CAAGGAGTACACCGCTCGTGGGAGGTTCCACCTTGTGATCCAGGAGAACCTCC 1307
Qy 1244 ACCCTGCTCTGCTGGGTGGGCAAGGCTACAGGCTGCTCTCTCTCTCTCTCTCTCTCTCT 1303

Db 1308 ACCCTGCTCTGCTGGGTGGGCAAAAGGCTACAGGCTGCTCTCTCTCTCTCTCTCTCTCT 1367
Qy 1304 CTGTCCCTCATGTGCTGCGCTGCTCCCGAGTACCTCGGAAAGTGGAGCATCGAGG 1363
Db 1368 CTGTCCCTCATGTGCTGCGCTGCTCCCGAGTACCTCGGAAAGTGGAGCATCGAGG 1427
Qy 1364 TAGGAGGAAACAGCAACCGGGGAGTCTCTGAGGCTGGGGCTGCCCTACCTCTACCCATT 1423
Db 1428 TAGGAGGAAACAGCAACCGGGGAGTCTCTGAGGCTGGGGCTGCCCTACCTCTACCCATT 1487
Qy 1424 CCCGACAGAGCTTGGCCCTTGGCTGCTGCCCTGCCCTGCTCTTTGGGGAACTGAGCTC 1483
Db 1488 CCCGACAGAGCTTGGCCCTTGGCTGCTGCCCTGCCCTGCTCTTTGGGGAACTGAGCTC 1547
Qy 1484 AGAGCGAGGTCTTTCAGAGAGGAAACAAATGAGGGTGGCAGGATATAAAGTCACT 1543
Db 1548 GGAGCGAGGTCTTTCAGAGAGGAAACAAATGAGGGTGGCAGGATATAAAGTCACT 1607
Qy 1544 CCATTCTCTACTCTCCATGCGAGCATGAACACAATTTCTCTCCACCTGGCTCCCAATTT 1602
Db 1608 CCATTCTCTACTCTCCATGCGAGCATGAACACAATTTCTCTCCACCTGGCTCCCAATTT 1666

RESULT 3
ABK40069
ID ABK40069 standard; DNA; 7924 BP.
XX AC ABK40069;
XX DT 21-MAY-2002 (first entry)
XX XX Human chemically pretreated gene sequence #76 strand 1.
DE Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytosine; ALDH6; CYP11A; CYP11B; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX OS Homo sapiens.
XX PN WO200202806-A2.
XX PD 10-JAN-2002.
XX PF 29-JUN-2001; 2001WO-EP07470.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-154757/20.
XX PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
XX useful for detecting cytosine methylation state of genes associated
XX with pharmacogenomics and for therapy of diseases e.g. cancer -
PS Claim 1; SEQ ID No 151; 24pp; English.
XX The invention relates to a nucleic acid comprising a sequence at
XX least 18 bases in length of a segment of the chemically pretreated DNA
XX of genes associated with pharmacogenomics according to one of the
XX sequences of the genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B
XX (NM_000497), CYP3A3 (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2
XX (NM_001979), OCLN (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360),
XX MRP (NM_004996, NM_019900, NM_019901, NM_019902, NM_019862, NM_019898,
XX NM_019899) and their complementary sequences, or a sequence (SI) chosen
XX from 87 sequences and their complements. The chemical pretreatment
XX is bisulphite treatment to convert cytosines (but not methyl-cytosines)
XX into uracils. Also included are an oligomer (II) in particular an
XX oligonucleotide or a peptide nucleic acid (PNA)-oligomer, comprising in

each case at least one base sequence having a length of 9 nucleotides which hybridises to or is identical to a chemically pretreated DNA of genes associated with pharmacogenomics and their complements, arranged in an array for analysing diseases associated with the methylation state (CpG) of the 87 sequences. The oligomers may also be used as PCR primers. The set of 87 nucleic acids and their complements is useful for diagnosis and therapy of solid tumours and cancer. The present sequence represents one the 87 DNA sequences or its complement. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

QY	1276	GGTGCTCTCTCTCCGCCACCCCTGCTCCCTCATGTGCCATGGCGCTGCCCTCCCC	1335
Db	1	GGTGTTTTTTTTTTTTTTTATTTATTTATGTTTTATGCTGTATGGTTTGT	60
QY	1336	AGTGACCTCGAAGTGGAGCATCGAGTAGGAGGAACACCAACCCGGAGTCCCTCGA	1395
Db	61	AGTGATTTTCGAAAGTGGAGTATCGAGGTAGGAGGAAACGGTAATTAGGAGTTTCGA	120
QY	1396	GCTTGGGCTGCTTACCTACCTACCTCCAAAGAGTGGGAAAGAACCAACCTTTC	1455
Db	121	GTGTGGGTTGTTTTATTTATTTATTTATTTTCGATTAGAGTTTGTGTTGTTCT	180
QY	1456	CCGCTCTCTCTTTGGGAACTGAGCTCAGAGCGAGTCTTCAGAGAAAGCAAAAT	1515
Db	181	TCGTTTGTGTTTTTGGGAAATGAGTTTATAGAGTGTGTTTATAGAGAAATTAAT	240
QY	1516	GAGGGTGGCAGGATTAAGGACCTCCATCTCTACCTCCCATGCGAGCATGAAACACA	1575
Db	241	GAGGGTGGTAGGATTAAGGAAAGTATTTATTTATTTATTTATTTATTTATTTAT	300
QY	1576	ATTCTCTCCACCTGCTCCCAATTAAGATGTGCAAGAGGCTGTGGGTACTCCAG	1635
Db	301	ATTTTTTTTATTTGTTTAAATTAAGATGTGCAATGAGTTTGTGGGTATTTAG	360
QY	1636	GGCAAGGAGAGCCCTGGGTCAGTGACACTGTTCAGCCCAACCATCTCCACAAAGG	1695
Db	361	GGTAAGGAGAGTTTGGGCTAGTCATATTTAGTTAAATTTATTTATTTATTTAT	420
QY	1696	GACATTTGGAATGAAGACTAGTCTCTATGATCAGGTTAAGCAAGGAGAGCTGG	1755
Db	421	GAGTATTTGGAATGAAGATTTAGTTTATTTAGTTAAGTTAAGGAGAGGTTGG	480
QY	1756	CCAGGACAGCAGTGTGCACAGCAGAGGGAATGTAGCAACAGCAGGCTCCTAGGCC	1815
Db	481	TTAGGATAGTACTGTTATAGTAGAGGGGAATCTAGTAATAGTAGGTTTATGGTT	540
QY	1816	CATCTTCCATTTCTAGGTAAAGAGCAATTTCCCTCAGACTCCCGCGGAGAGCTAGC	1875
Db	541	TATTTTATTTTATTTAGTAAAGAGATTTTATTTTATTTTATTTTATTTTATTT	600
QY	1876	CTAGCCTTCAGCAACCAAGTTCTCTCGGACCCCAAGTTTATGGAGAGGCAAGAC	1935
Db	601	TTAGTTTATTTATTTAGTAAAGGTTTTTTTGGGATTTAAAGTTTATGGAGAGG	660
QY	1936	TTTCATGGAGAGAGAGGAGCCCTGGTGAAGAGCTGTGCTGTCTCTCTTGGCC	1995
Db	661	TTTATGGGAAGAGAGAGGAGGTTTTTGGGTAGAAACGTTTGTGTTTGTGTTT	720
QY	1996	TTTAAAGCAAGCGCTCATCTTCCCTCTACCTCTCTAGAGGCTTGGAGGTTTCCCAAC	2055
Db	721	TTTAAAGATAAAGCGTTTATTTGTTTATTTTATTTTATTTTATTTTATTTTAT	780
QY	2056	ACACTGTGGCTACAGGTGGAGGGAAGAGGACTCCTCTCCAGAGTGTCTATGTCAGGA	2115

Db	781	ATATTGTGTTATAGCTGAGGGAAGAGGATTTTTTTTTTATAGAGTGTATTTTAGGAA	840
QY	2116	GTTCCTTTAAACCCCATATGCGCCCAAGAGTAGCTCTAGAGGCCCTTTAAACACGGAACA	2175
Db	841	GTTTTTTTTAAATTTTATATGTTTAAAGAGTAGTCTGAGGAGGTTTTTAAACACGGAATA	900
QY	2176	AGTAATTTACCAGTTCTACTGGGTTCTGCCACCGTCCCAAGGTGGCGAGGCTAGG	2235
Db	901	AGTAATTTATTTAGTTTATTTGGGTTTTTGTATCGTTTTTAAGGTGGCGAGGTTAGG	960
QY	2236	AAGAGGTCATTTTAAAGCCACACATTTAGCTGCACCTGCTGCTGCAGCCCAACAAAGA	2295
Db	961	AAAGAGGTTATTTTAAAGTTATATATTAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1020
QY	2296	ACTGCTGTGTGAGTATTCATCACTAAAGAACCAACCAACCAACCACTCATATGTGAAGA	2355
Db	1021	ATTGGGTGTGTGAGTATTTATTAATTAAGAAATTAAGAAATTTAGGATTTATATGTGAGA	1080
QY	2356	TAAGAACCTCACTTCTTACTCTCCAAAGAGTGGGAAAGAACCAACCAACCTTTC	2415
Db	1081	TAAGAAATTTTATTTTATTTTATTTTAAAGAGTGGGAAAGAAATTTATTAATTTTTT	1140
QY	2416	CTCTGACTTACCAACACAGGAAACACAGGAGAGGTTGGCTCAGGACTTAGGACAGG	2475
Db	1141	TTTTTGTATTTTAAATTTAGGAAATTAGTAGGAGGTTGGTTTAGGATTTAGGATAGG	1200
QY	2476	GTATAGCTTAGATGTGGAAAGCAAGAGAGCAGGAGAACTTGTAAATCACTGCTCAATGA	2535
Db	1201	GTATAGTTTATAGTGTGGAAAGTAAAGAGAGTAGGAACTTCTAAATTTATGTTAATCA	1260
QY	2536	GAAGAGGAGACAGCTAATCTAGGATGAAGCTGTACTAGGCTGGAGTTGCTTCTTGAA	2595
Db	1261	GAAGAGGAGATAGTTAAATTTAGGATGAAGTTCTGATAGTTGGAGTTGTTTTTTGAA	1320
QY	2596	GATGGGACTCTTGGGTATCAAGACCTATGCCACATCACACTGGGCTAGGGAAGTAGCT	2655
Db	1321	GATGGATTTTGGGTATTAAGATTTATGTTATATTTATTTGGGTTAGGGAAGTAGGT	1380
QY	2656	GATGCCAGCCCTCAAGCTGTCTTTCAGCCAGGAGCTTGAGAAGTTATATTTGGSCAGTGC	2715
Db	1381	GATGTTAGTTTAAAGTTTGTGTTTATAGTTAGGATTTGAGAAGTTATATTTGGGTAGCT	1440
QY	2716	TCCATCTCTGACACAGTATTTTCCCTTCCCTGGAAGATCAGCAGGCTGCTCATTTG	2775
Db	1441	TTTAAATTTCTGATTTAGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT	1500
QY	2776	TCTTCTCTCTAGCCCTCAGGAAAGAGGACTATATTTGTACTGTACCTTAGGGTT	2835
Db	1501	TTTTTTTTTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	1560
QY	2836	CTGGAAGGGAACATGGAATCAGGATTTATAGACTGTATAGGCTTATCCCAAGGGCC	2895
Db	1561	TTGGAAGGGAACATAGGAATTTAGGATTTATAGATTTAGGTTTATTTATTAAGGGTT	1620
QY	2896	ATGACTGGGAAAGGATGGGACCAAGAGAAATTTGGGATTTTATTTTATTTTATTTTATTT	2955
Db	1621	ATGATTTGGGAAAGGATGGGAGTAGAAGAGAAATTTGGGATTTTATTTTATTTTATTTT	1680
QY	2956	CACCTTAAACTTTTGTGGCTGGGCTGTCTGTAGGCCACAGCTGTTAAGCAGGCTCT	3015
Db	1681	TATTTTAAATTTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTT	1740
QY	3016	GCTGGCTGTTTACTCGTCACCACTCTGCACCTCTGCACCTGCTGTCTTGAGACTCCATCCAG	3075
Db	1741	GTGGTTTGTGTTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT	1800
QY	3076	AGCAGCCACCTGCTGCTGAGCCTCCACTATCTCCTCTGAGCGGCTGAGCTTCTGCTGAC	3135
Db	1801	AGTAGCTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT	1860
QY	3136	TGCTGCTGGGCTCATATATGAAATTTGAGCAGGTTTCATCTATTTTAAACACAGATGTT	3195

Db 1861 TGTGTTCCGGTTTATATATGAAATGTGTAGTAGGCTTTATTTATTTAAATATATGATGTT 1920
QY 3196 TACAAATAAGATTTATTTCAACACC 3223
Db 1921 TATAAATAAGATTTATTTAAATATC 1948
RESULT 4
ID ABL34138
XX ABL34138 standard; DNA; 7924 BP.
XX ABL34138;
XX 26-MAR-2002 (first entry)
XX Human immune system associated gene SEQ ID NO: 2111.
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosolic; neurotropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX Homo sapiens.
XX OS
XX WO200200928-A2.
XX PN
XX 03-JAN-2002.
XX PD
XX 02-JUL-2001; 2001WO-EP07537.
XX PF
XX 30-JUN-2000; 2000DE-1032529.
XX PR
XX 01-SEP-2000; 2000DE-1043826.
XX PA
XX (EPIC-) EPIGENOMICS AG.
XX PI
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX cytosine methylation
XX
XX Claim 1; SEQ ID NO 2111; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
XX Sequence 7924 BP; 2087 A; 83 C; 2037 G; 3717 T; 0 other;
SQ
Query Match 36.6%; Score 1192.8; DB 24; Length 7924;
Best Local Similarity 75.8%; Pred. No. 5e-279;
Matches 1476; Conservative 0; Mismatches 472; Indels 0; Gaps 0;
QY 1276 GGTGCTTCTCTTCCCCACCCCACTGTCCTCATGTGCCATGCGCTCCCTCC 1335
Db 1 GGTGCTTCTCTTCCCCACCCCACTGTCCTCATGTGCCATGCGCTCCCTCC 60
QY 1336 AGTGACCTCGGAAGTGGAGATCCAGGTAGGAGGGAACACACCGGGAGTCTCGA 1395
Db 61 AGTGATTTGCGAAGTGGAGTATAGGATAGGAGGGAACCGGTAATAGGAGTTTCGA 120
QY 1396 GCCTGGGCTGCCCTACCTCTACCCATTCCCGACGAGCTTTGCCCTTGCTGGCTGC 1455

Db 121 GTTTGGGTGTTGTTTATTTTATTTTATTTTTCGATTAGAGTTTGTGTTGTTGTTGT 180
QY 1456 CGCCCTGCTCTTTTGGGGNACTGAGCTCAGAGCGAGTGCCTTCAGAGAAGGAACAAT 1515
Db 181 TCGTTTGTGTTTTTTGGGGAATTTAGAGGTAGGTGTTTGTAGAGAAGGAATAAAT 240
QY 1516 GAGGGTGGCAGGATATAAAGTCACCTCCATTCCTACCTCCCATGACATGAACACA 1575
Db 241 GAGGGTGGTAGGGATATAAAGTATTTTATTTTATTTTATGTAATGAATATA 300
QY 1576 ATTTCTCTCCACTGCTGCCCAATTTAAAGATGGAGCAAGGCTGTGTTACTCCAG 1635
Db 301 ATTTTATTTTATTTGTTTAAATTTAAAGATGCGATTAAAGTTTGTGGGTATTTAG 360
QY 1636 GGGCAAGGAGAGCCCTGGGTCAGTCACACTCTCAGGCCAACCATGCCATCCACAAGG 1695
Db 361 GGGTAAGGAGAGTTTGGGTTAGTGATATTTAGTTAAATATATGTTATTAAGGG 420
QY 1696 GAGCATTTGGAATGAAGGACTAGCTCCTATGATATCAGGTTAAGAGCAAGGAGAGCTGG 1755
Db 421 GAGTATTTGGAATGAAGGATTTAGTTTATGATTTAGTTAAGAGTAAGGAGAGTGG 480
QY 1756 CCAGGACAGCAGTTTTCACAGCAGAGGGGAATGTAGCAACAGCAGGGCTCTCTAGGCC 1815
Db 481 TTAGGGATAGTAGTTTGTATATAGTAGGGGAATGTAGTAATAGTAGGGTTTGTAGGTT 540
QY 1816 CATCTTCATTTCTTAGTGAAGAGCATTTCTCAGACTCCAGGCGGAGAGACTGAGC 1875
Db 541 TATTTTATTTTATTTTAGTTAAGAGATTTTATTTTATTTTATTTAGTTTGTAGGTTGAGT 600
QY 1876 CTAGCCTTCAGCAACCAAGGTTCTCTGGGACCCAAAGTTTATGGGAGAGGCAAGAC 1935
Db 601 TTAGTTTATTTAGTAATTAAGGTTTGTGGATTTAAAGTTTATGGGAGGGTAAGAT 660
QY 1936 TTCTATGGGAAGAGAGGAAGGCCCTGGGTAGAACCGTTGGTGTCTCTCTCTTGGCC 1995
Db 661 TTTATGGGAAGAGAGGAAGGAGTTTGGGTAGAACGTTTGGTGTGTTTGTGTTTGGTT 720
QY 1996 TTTAAGACAAAGCGTCTATCTTGGCTCTACCTCTGATAGGCTTGAGGGTTGCCAAC 2055
Db 721 TTTAAGATAAAGCGTTTATTTTGTGTTTATTTTATTTTATTTTATTTTATTTTATTT 780
QY 2056 ACACGTGTGCTACAGTGGAGGAAGAGGACTCTTCTCCAGAGTGCATGTTTCAGGA 2115
Db 781 ATATTGTGGTTATAGTGGAGGAAGAGGATTTTATTTTATTTTATTTTATTTTATTTAGGA 840
QY 2116 GTTCTTTAAACCCCATATGGCCCAAGAGTAGCTCGTAGGAGGCCCTTTTAAAGACGGAACA 2175
Db 841 GTTTTATTTTAAATTTATATGTTTAAAGTAGTTCGTAGGAGGTTTATTTTAAAGCGGAATA 900
QY 2176 AGTAATTTTACCAGTTCTACTGGGTTCTGCCCCACCGTCCCAAGGTGGCGGAGGCTTAG 2235
Db 901 AGTAATTTTATTTAGTTTATTTTGGGTTTGTGTTTATTTTATTTTAAAGTGGCGAGGTTAG 960
QY 2236 AAGAGGGTCATCTCTTAAAGCCACATATAGCTGCCTGCTGCTGAGCAACCAACGAAGA 2295
Db 961 AAGAGGGTTATTTTAAAGTTATATATAGTTATTTGCTGTTGTTGTTTAAATATAAGA 1020
QY 2296 ACTGGGTGTGAGTATTTCACTCACTAAGAACCAAAATCCAGGGCTCATATGTGAAGA 2355
Db 1021 ATTTGGGTGTGAGTATTTATTTAATTAAGAAATTTAAATTTTATTTTATTTTATTTAGGA 1080
QY 2356 TAAGAACCTCACTCTCTTACTCTCCAAAGAGTGGGGAAGAACCAATCAACCTTTC 2415
Db 1081 TAAGAAATTTATTTTATTTTATTTTAAAGAGTGGGGAAGAAATTTTAAATTTT 1140
QY 2416 CTCCTGACTTACCAACAGGAAACAGCAGAGGGGTGGCTCAGGACTTAGGACAG 2475
Db 1141 TTTTGTATTTATTAATTTAGGAAATATAGTAGGAGGGGTGGTTTATTTTATTTAGGATAG 1200
QY 2476 GTATAGCTTAGATGTTGGAAGCAAGGAGAGCAGGAGCTTCTAAATCACTGCTGCTAAATCA 2535

Db 1201 GTATAGTTTATAGTGTGGAAGTAAAGAGAGTAGGAACTTCTAAATATTGTTTAATGA 1260
QY 2536 GAAAGGAGACAGCTAACTAGGATGAAGCTGTGACTAGGCTGGAGTTGCTTCTTGAA 2595
Db 1261 GAAAGGAGATAGTTAAATTTTAGATGAAGTGTGATTAAGTTGGAGTTGTTTTTGA 1320
QY 2596 GATGGGACTCCTTGGGTATCAAGACCTATGCCACATFACACACTGGGGCTAGGGAAGTAGT 2655
Db 1321 GATGGGATTTTGGGTATTAAGATTATGTTATATATATATTGGGCTTAGGGAAGTAGT 1380
QY 2656 GATGCCAGCCCTCAAGTCGTCTTACCCAGGAGCTTGAGAAGTTATATGGCAGTGCC 2715
Db 1381 GATGTTAGTTTAAAGTTGTTTTTATAGTTAGGAGTTGAGAAGTTATATGGGTAGTGT 1440
QY 2716 TCCAAATCTGGACCACTATTTCCAGCTTCCCTGGAAGATCAGCAGGGTGCCATTCATG 2775
Db 1441 TTTAAATTTGGGATTTAGTATTTTATTTAGTTTATTTTGAAGATTAGTAGGGTTATTTATG 1500
QY 2776 TCTTTCTCTCTAGCCCCCTCAGGAAGAGGACTATATTTTGTACTTACCCCTAGGGGTT 2835
Db 1501 TTTTTTTTTTATGTTTTTTAGGAAAGAGGATTATATTTGTATGTATTTTAGGGGTT 1560
QY 2836 CTGGAAGGAAAACATGAATCAGGATCTATAGACTGATAGCCCTATCCCAAGGGCC 2895
Db 1561 TTGGAAGGAAAATATGGAATTAGGATTTTATAGATTGATAGTTTATTTATAGGGTT 1620
QY 2896 ATGACTGGGAAAAGGTATGGAGCAGAGAGAAATGGGATTTTAGGGTGCAGCTACGCT 2955
Db 1621 ATGATTTGGGAAAAGGTATGGAGTAGAGAGAAATGGGATTTTAGGGTGTAGTGT 1680
QY 2956 CACCTAAACTTTTGGGCTGGGCAATCTTTGAGGCCAGACGTGTTAAGCAGGCTCT 3015
Db 1681 TATTTTAAATTTTGGTGGTTGGGTATCTTTTGGAGTTTAGATTGTTAAATAGGTTT 1740
QY 3016 GCTGGCTGTCTACTCTCACCCTCTGCACCTCTGCTGTGAGCTCCATCCAGCC 3075
Db 1741 GTTGGTTGTTTATTCCTTATATTTTGTGATTTGTTGTTGATATTTATTTAGTTT 1800
QY 3076 AGGCACGCCACCTGCTCCTGAGCCTCCACATCTCCCTGTGACGGGTGAACCTGCTAC 3135
Db 1801 AGGTACGTATTTGTTTGGTGTATTTATTTATTTTGTGACGGGTGAATTCGCTGTAT 1860
QY 3136 TGTGTCGCGGTCCATATATGAATTTGAGCAGGGTTCACTATTTTAAACACAGATGTT 3195
Db 1861 TGTGTTTCGGGTTATATATGAATTTGTGAGTAGGGTTTATTTATTTAAATATAGATGTT 1920
QY 3196 TACAAAATAAAGATTATTTCAAAACCACC 3223
Db 1921 TATAAATAAAGATTATTTTAAATATC 1948

RESULT 5
ABK40070/c
ID ABK40070 standard; DNA; 7924 BP.
XX AC ABK40070;
XX DT 21-MAY-2002 (first entry)
XX DE Human chemically pretreated gene sequence #76 strand 2.
XX KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
XX KW cytosatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
XX KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX OS Homo sapiens.
XX PN WO200202806-A2.
XX PD 10-JAN-2002.
XX PF 29-JUN-2001; 2001WO-EP07470.
XX

PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-154757/20.
DR
XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated
PT with pharmacogenomics and for therapy of diseases e.g. cancer -
XX
PS Claim 1; SEQ ID No 152; 24pp; English.
XX The invention relates to a nucleic acid comprising a sequence at
CC least 18 bases in length of a segment of the chemically pretreated DNA
CC of genes associated with pharmacogenomics according to one of the
CC sequences of the genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1
CC (NM_000497), CYP3A3 (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2
CC (NM_001979), OCLN (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360),
CC MRP (NM_004996, NM_019900, NM_019901, NM_019902, NM_019862, NM_019898,
CC NM_019899) and their complementary sequences, or a sequence (S1) chosen
CC from 87 sequences and their complements. The chemical pretreatment
CC is bisulphite treatment to convert cytosines (but not methyl-cytosines)
CC into uracils. Also included are an oligomer (II) in particular an
CC oligonucleotide or a peptide nucleic acid (PNA)-oligomer, comprising in
CC each case at least one base sequence having a length of 9 nucleotides
CC which hybridises to or is identical to a chemically pretreated DNA of
CC genes associated with pharmacogenomics and their complements, arranged in
CC an array for analysing diseases associated with the methylation state
CC (CpG) and/or detecting SNPs (single nucleotide polymorphisms)
CC of the 87 sequences. The oligomers may also be used as PCR primers.
CC The set of 87 nucleic acids and their complements is useful for diagnosis
CC and therapy of solid tumours and cancer. The present sequence
CC represents one the 87 DNA sequences or its complement.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7924 BP; 1973 A; 83 C; 1827 G; 4041 T; 0 other;

Query Match 35.6%; Score 1160.8; DB 24; Length 7924;
Best Local Similarity 74.1%; Pred. No. 2,9e-271;
Matches 1466; Conservative 0; Mismatches 512; Indels 0; Gaps 0;

QY 1278 TGCTTCTCTCTTCCCTCCACCCCACTGTCCCTCATGTGCGCATGGCGCTGCCCTCCACG 1337
Db 7922 TACTTCTCTCTTCCCTCCACCCCACTATCCCTCATATACCATAAACCTACCTCCCAA 7863
QY 1338 TGACCTCGAAAGTGGAGCATCGAGGTAGGAGGAAACACACACCGGGAGTCTCGAGC 1397
Db 7862 TAACCTACGAAATAAACAATCGAATAAATAAATAAACAACCAACCAAAATCTCTCGAAC 7803
QY 1398 CTGGGGCTGCCCTTACCTCTACCCATTCCTCCGACAGAGCTTTGCTCTTGGCTGCC 1457
Db 7802 CTAAACTACCTTACCTCTACCCATTCCTCCGACCAAACTTTACCTTACTTACTACCC 7743
QY 1458 GCCTGCTCTTTGGGGAACCTCAGCTCAGAGCGAGGTGCTTCAGAGAAGGAAACAAATGA 1517
Db 7742 GCCTTACCTCTTTAAATAAACTAAACTCAAAACAAATACTTCAAAAAAATAAATAA 7693
QY 1518 GGGGTGCGAGGATTAAGTACCTCCATCTCTACTCCATGCGAGCATGAACACAAT 1577
Db 7682 AAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7623
QY 1578 TTCTCTCCACTGCTGCCAAATTTAAAGATGTGAGCAAGGCTGTGGGTACTCCAGG 1637
Db 7622 TTCTCTCCACTTAACTCCAAATTTAAATAAATAAATAAATAAATAAATAAATAA 7563
QY 1638 GCAAGGAGAGCCCTGGGTGCTAGTGTGACACTGTGACGCCACCATGCATCCCAAGGGA 1697
Db 1638 GCAAGGAGAGCCCTGGGTGCTAGTGTGACACTGTGACGCCACCATGCATCCCAAGGGA 1697

Db 7562 ACAAAAAAAAAAACCCTAAATCAATACACTATCAAAACCAACCATCACTCCACAAAAA 7503
QY 1698 GCATTGGAAATGAAGGACTAGCTCCTATGATCAGGTTTAAGACGAAGGAGCTGGCC 1757
Db 7502 ACATTTAAAAATAAAAACTACTCCTATATCAATTAATAAACAACAAAAAATCAACC 7443
QY 1758 AGGACAGAGTTTGCACAGAGAGGGGATATAGACAGCAGGCGCTTCTAGGCCCA 1817
Db 7442 AAAAAACAATTTACACAACAAAAAATAATACAACAACAAAAACCTCTTAAACCCCA 7383
QY 1818 TCTTCCATTTCTAGTATAGAGACATTTCTCAGACTCCAGCGGAGGACTGAGCT 1877
Db 7382 TCTTCCATTTCTAATAAAAAAATTTCTCCTCAAACTCCCAACGAAAACTAAACCT 7323
QY 1878 AGCCTTCAGCAACCAAGGTTCTCCCTGGGACCCAAAGTTTATGGGAGAGGGCAAGACTT 1937
Db 7322 AACCTTCACACCAAAATTTCTCTTAARACCCAAATTTATAAAAAACAACAAACTT 7263
QY 1938 CATGGAAGAGAGAAGGAGGCGCTGGGTAGAAACGCTTGTGCTGTTCTCTTTGGCCCT 1997
Db 7262 CATAAAAAATAAAAAAACCCTAAATAAAAAACGCTTAATACTATTCTCTTTAACCTT 7203
QY 1998 TAAGCAAAAGCCTCATCTTGGCCCTCTAGCTCCTGATAGGCTTGGGGTTTCCCAACAC 2057
Db 7202 TAAAAACAACGCTCATCTTACCCCTCTACCTCTAATAAATTAATAAATTTTACCAACCCAC 7143
QY 2058 ACTGTGGCTACAGGTGGAGGAGGAGGACTCCTCTCCAGAGTGTATGTTTCAGGAAGT 2117
Db 7142 ACTATAACTACAATAAATAAAAAAANAACCTCTCTCCAAATACTATATTCAAAAT 7083
QY 2118 TTTCTTAACCCCATATGGCCCAAGAGTAGCTGTAGGAGGCGCTTTAAAGACGGAACAG 2177
Db 7082 TTTCTTTAACCCCATATAACCCCAAAATAACTCTGTAATAAACCTTTTAAAAACGCAACAA 7023
QY 2178 TAATTTACAGTTCTACTGGGTTCTGCCACCGTCCCAAGGTGGCGGAGGCTAGGAA 2237
Db 7022 TAATTTACCAATTTACTATAAATTTCTACCCACCGTCCCAAAATAAACGAACTTAAAAA 6963
QY 2238 GAGGCTCACTTTAAAGCCACACATTAAGCTGCACTGCTGGTGCAGCCAAAAACAAGAAC 2297
Db 6962 AAAAAATCACTTTAAACACACATTAACCTACACTAGTAACTACAACCAACAAACAAAC 6903
QY 2298 TGGGTGTTGAGTATCAATCAAGAACCAAAATCCAGGGCACTCATATGTGAAGGATA 2357
Db 6902 TAAATATTAAATATTCACTAAACCAAAATCCAAACACTCATATATAAAAAATA 6843
QY 2358 AGAACCTCACTTCTTACTCTCCAAAGAGAGTGGGGAAGCAACCAACCTTTCT 2417
Db 6842 AAAAACTCACTTCTTACTCTCCAAAAAAAATAAAAAAACAACCAACCTTTCT 6783
QY 2418 CCTGACTTACCAAAACAGGAAACAGCAGGAGGCTGCTCAGGACTTAGGACAGGCT 2477
Db 6782 CTTAATTTACCAACCAAAACCAACCAAAATAAATACTCAAACTTAAAAACAAT 6723
QY 2478 ATAGCTTAGATGGTGGAAAGCAAGAGAGAGAGTGTGTAATCACTGGCTAATAGA 2537
Db 6722 ATAACCTTAAATATAAAAAACAACAAAAAATAAAAAATTAATACTAATAAAAA 6663
QY 2538 AAAGGACAGACTAATCTAGGATGAGCTGACTAGCTGGAGTGGCTTCTTGAAGA 2597
Db 6662 AAAAAACAACACTAATCTTAAATAAATAAATACTAATACTAATACTTCTTAAAAA 6603
QY 2598 TGGGACTCTTGGGTATCAAGACCTATGCCACATCACACTGGGCTAGGGAAGTAGGTA 2657
Db 6602 TAAACTCTTAAATATCAAACTTATACCATATACACTAATACTAATAAATAAATAA 6543
QY 2658 TGCAGCCCTCAAGTCTGTTTACGCCAGGACTTGAAGAGTTATATTGGGAGTGGCTC 2717
Db 6542 TACCAACCTCAATATCTCTTCAACCAAAACCTTAAAAAATATATTAAACAATACTC 6483
QY 2718 CAATCTGGGACAGTATTTTCACTTCTCCGAAAGTCAAGGAGGCTCATTTGTC 2777
Db 6482 CAATCTAATAACCAATATTCTCACTTCTCCCTAAAAATCAACAAAAATACCATTCATTATC 6423

QY 2778 TTTCTCTCTAGCCCTCAGGAAAGAGGACTATATTTGTACTGTACCTTAGGGTTCT 2837
Db 6422 TTTCTCTCTACCCCTCAGAAAAAATACTATATTTATATATATATATATATATCT 6363
QY 2838 GGAAGGAAAAACATGGAATCAGATTTCTATAGCTATAGGCGCTTATCACAAGGGCCAT 2897
Db 6362 AAAAAAATAAATAAATACTTCTATAAATCTATAAATCTATAAATCTATCCAAAAACCAT 6303
QY 2898 GACTGGAAAAAGTATGGGAGCAGAGGAATTTGGGATTTTGGGTGAGCTACGCTCA 2957
Db 6302 AACTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6243
QY 2958 CCTTAACTTTTGGTGGCTGGGCGATGCTTTGAGGCCAGACTGTGTAAGCAGGCTCTGC 3017
Db 6242 CCTTAACTTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 6183
QY 3018 TGGCTGTTTACTGTGACCACTCTGACCTGTGCTGTGAGACTTCTATCAGGCGCCAG 3077
Db 6182 TAACTTATTACTGTGACCACTCTGACCTGTGCTGTGAGACTTCTATCCTTAACTTAACTTAACT 6123
QY 3078 GCACGCCACTCTGCTGAGCTCCACTATCTCCCTGTGACGGTGAACCTTCTGTACTG 3137
Db 6122 ACACGCCACTCTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 6063
QY 3138 TGTCTCGGTCCATATATGATTTGTGACGAGGCTTCATCTATTTTAAACACAGATGTTTA 3197
Db 6062 TATCTCGATCCATATATAAATTAATAAACAATAAATTCATCTATTTTAAACACAAATATTTA 6003
QY 3198 CAAAATAAAGATTTTCAACACCAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 3257
Db 6002 CAAAATAAATAAATTTCAACACCAAGATATAACTACCTTAAATAAATAAATAAATAAATAAATAA 5943

RESULT 6

ABL34139/c
ID ABL34139 standard; DNA; 7924 BP.
XX AC ABL34139;
XX AC ABL34139;
DT 26-MAR-2002 (first entry)
XX Human immune system associated gene SEQ ID NO: 2112.
DE Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosinatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antirheumatic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX Homo sapiens.
OS Homo sapiens.
PN WO200200928-A2.
XX WO200200928-A2.
PD 03-JAN-2002.
XX 02-JUL-2001; 2001WO-EP07537.
PF 02-JUL-2001; 2001WO-EP07537.
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
PI WPI; 2002-130909/17.
DR Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -

Qy	3198	CAAAATAAGATTATTTCAAACCAACCACCAAAAAAAAAAAAAAAAAAAAAA 3257
Db	6002	CAAAATAAAATTTATTTCAAACCAACCGATATACCTTAATAATCCTTAAAAATPAAA 5943

PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465460/50.
DR P-PSDB; AAU17136.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders
XX
PS Claim 1; SEQ ID No 88; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
CC pathway protein coding sequences and PCR primers of the invention.
XX
Query Match 35.6%; Score 1160.6; DB 22; Length 2021;
Best Local Similarity 99.2%; Pred. No. 1.9e-271;
Matches 1166; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 2 GCCGGCTGCCGGCCGCGAGTTCGGGGCCCGCTGGCCCGCCAGTCATGCGGAGCAGTACGA 61
DB 38 GCCGGCTGCCGGCCGCGAGTTCGGGGCCCGCTGGCCCGCCAGTCATGCGGAGCAGTACGA 97
QY 62 TGTGCTCTTCCGGCTGCTGCTGATCGGGGACTCCGGGGTGGGAAGACCTGCCTGCTGTG 121
DB 98 TGTGCTCTTCCGGCTGCTGCTGATCGGGGACTCCGGGGTGGGAAGACCTGCCTGCTGTG 157

QY 122 CCGCTTACCGACACGAGTTCACACTCCCTCGCACATCTCCACCATCGTGTGACTTTAA 181
DB 158 CCGCTTACCGACACGAGTTCACACTCCCTCGCACATCTCCACCATCGTGTGACTTTAA 217
QY 182 GATGAAGACCATAGAGGTAGCGGCATCAAAAGTCCGGATACAGATCTGGGACACTGCAGG 241
DB 218 GATGAAGACCATAGAGGTAGCGGCATCAAAAGTCCGGATACAGATCTGGGACACTGCAGG 277
QY 242 GCAGGAGAGATACACGACCATCACAAAGCAGTACTATCGCGGGGCCAGGGATATTTT 301
DB 278 GCAGGAGAGATACACGACCATCACAAAGCAGTACTATCGCGGGGCCAGGGATATTTT 337
QY 302 GGTCTATGACATTAGCAGCGAGCGCTTTACCAGCACATCATGAAGTGGGTGAGTGCACGT 361
DB 338 GGTCTATGACATTAGCAGCGAGCGCTTTACCAGCACATCATGAAGTGGGTGAGTGCACGT 397
QY 362 GATGAGTAGGACACAGAGGGGTCCAGAAAGTCTTATTTGGGAATAAGCTGATGAGGA 421
DB 398 GATGAGTAGGACACAGAGGGGTCCAGAAAGTCTTATTTGGGAATAAGCTGATGAGGA 457
QY 422 GCAGAAACGCGAGGTGGGAAGAGAGCAAGGCGAGCAGCTGGCGAAGAGTATGGCATGGA 481
DB 458 GCAGAAACGCGAGGTGGGAAGAGAGCAAGGCGAGCAGCTGGCGAAGAGTATGGCATGGA 517
QY 482 CTTCTATGAACAAGTGCCTGCACCAACCTCAACATTAAGAGTCAATTCACGCGTCTGAC 541
DB 518 CTTCTATGAACAAGTGCCTGCACCAACCTCAACATTAAGAGTCAATTCACGCGTCTGAC 577
QY 542 AGAGCTGGTGTCCAGGGCCCATAGGAAGAGCTGGGAAGGCTCCGGATGGTGGCCAGCAA 601
DB 578 AGAGCTGGTGTCCAGGGCCCATAGGAAGAGCTGGGAAGGCTCCGGATGGTGGCCAGCAA 637
QY 602 TGAGTTGGCAGCTGGCAGAGCTGGAGGAGGAGGAGGCAAAACCCGAGGGCCAGCGAACTC 661
DB 638 TGAGTTGGCAGCTGGCAGAGCTGGAGGAGGAGGAGGAGGCAAAACCCGAGGGCCAGCGAACTC 697
QY 662 TTGCAAAACCTGCTGGTGTGAGTCTCTGTGTGGGGCACCCACACGACACCCCTCTTCCC 721
DB 698 TTGCAAAACCTGCTGGTGTGAGTCTCTGTGTGGGGCACCCACACGACACCCCTCTTCCC 757
QY 722 TCAGGAGGCGCTGGCAGACAGCGGGGCTTTGCCCTGCTGCTGCTCTCTGCTGCTG 781
DB 758 TCAGGAGGCGCTGGCAGACAGCGGGGCTTTGCCCTGCTGCTGCTGCTCTCTGCTG 817
QY 782 TGATGACCTTATGAGTATCAGTAGCCACTACTCCCTGCTGCTGCTGCTGCTGCTG 841
DB 818 TGATGACCTTATGAGTATCAGTAGCCACTACTCCCTGCTGCTGCTGCTGCTGCTG 877
QY 842 TGCTGTCTATCTAAGCAGCCCTGTCCCGAGCCGCTCCACCCCTGGAGTGTCTTCTTCAG 901
DB 878 TGCTGTCTATCTAAGCAGCCCTGTCCCGAGCCGCTCCACCCCTGGAGTGTCTTCTTCAG 937
QY 902 CCGTGTTCCTCCAGCAGCGCTGTACGACCCCGCAGGATGTCCGCAAGCACTGTCTCA 961
DB 938 CCGTGTTCCTCCAGCAGCGCTGTACGACCCCGCAGGATGTCCGCAAGCACTGTCTCA 997
QY 962 CCATCCCGCAGCCAGCAGACAGCGGGCTGGAGTCCAGGCCACTTTTCAGTGTCTC 1021
DB 998 CCATCCCGCAGCCAGCAGACAGCGGGCTGGAGTCCAGGCCACTTTTCAGTGTCTC 1057
QY 1022 TTTCTCGTGCATCGTGTCTCTCTGCTGCTTTTCTCTCTTCTCTCTCTCTCTCTCTCT 1081
DB 1058 TTTCTCGTGCATCGTGTCTCTCTCTGCTGCTTTTCTCTCTCTCTCTCTCTCTCTCTCT 1117
QY 1082 GACCCCTCCCTCCGGTGCCTTTTCGTATCAAAAGTCTCTCAAAACCCCGCTCCCGTGTGTC 1141
DB 1118 GACCCCTCCCTCCGGTGCCTTTTCGTATCAAAAGTCTCTCAAAACCCCGCTCCCGTGTGTC 1177
QY 1142 CTGCTGTGTCAGCTCGCTCTTTCTCTTCTTCTCTTCTCTTCTCTTCTCTTCTCTCT 1176
DB 1178 CTGCTGTGTCAGCTCGCTCTTTCTCTTCTTCTCTTCTCTTCTCTTCTCTTCTCTCT 1212

RESULT 8
AAZ24400
ID AAZ24400 standard; cDNA; 895 BP.
AC AAZ24400;
XX
DT 14-FEB-2000 (first entry)
XX
DE Human bladder tumour cDNA library derived EST 12.
XX
KW Expressed sequence tag; human; bladder; tumour; cancer; cytostatic;
KW treatment; gene therapy; EST; ss.
XX
OS Homo sapiens.
XX
PN DE19818619-A1.
XX
PD 28-OCT-1999.
XX
PF 21-APR-1998; 98DE-1018619.
XX
PR 21-APR-1998; 98DE-1018619.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;
XX WPI; 1999-612028/53.
XX
DR New nucleic acid sequences expressed in bladder tumor tissue, and
XX derived polypeptides, for treatment of bladder tumor and identification
XX of therapeutic agents -
XX
PS Claim 3; Page 69; 132pp; German.
XX
CC This invention describes novel polypeptide fragments (I) and the
CC polynucleotides (II) that encode them that are highly expressed in a
CC human bladder tumour and which have cytostatic activity. (II) are used
CC for recombinant expression of (I) and to isolate complete genes. (I) are
CC used to identify agents suitable for treatment of bladder cancer, to
CC directly treat this form of cancer (including expression from gene
CC therapy vectors) or are used in a preparation for cancer treatment. (I)
CC is also used for the generation of specific antibodies. (II) are
CC identified by assembling ESTs (expressed sequence tags) from a
CC particular tissue type before comparison of expression patterns. This
CC allows a significantly longer fragment of the gene to be revealed, and
CC therefore reduces the number of failures associated with the fact that
CC ESTs from different libraries may represent different parts of the same
CC unknown gene, distorting the estimated frequency of occurrence in a
CC particular tissue. AAZ43260-243309 represent expressed sequence tag (EST)
CC fragments isolated from a human bladder tumour cDNA library which encode
CC the proteins represented in AAZ66143-Y66198.
XX
SQ Sequence 895 BP; 249 A; 205 C; 228 G; 213 T; 0 other;

Query Match 26.6%; Score 866.2; DB 20; Length 895;
Best Local Similarity 99.5%; Pred. No. 3.8e-200;
Matches 879; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 2358 AGAAGCTTACTTCTTACTCTCCAAAAGAGAGTGGGGAAGAACATCAACCTTCTCT 2417
DB 14 AGAAGCTTACTTCTTACTCTCCAAAAGAGAGTGGGGAAGAACATCAACCTTCTCT 73
QY 2418 CTGAGCTTACCAACAGCAAGAACAGAGAGAGTGGCTCAGGACATTTAGGACAGGT 2477
DB 74 CTGAGCTTACCAACAGCAAGAACAGAGAGAGTGGCTCAGGACATTTAGGACAGGT 133
QY 2478 ATAGCTTAGTGGTGAAGCAAGAGAGAGAGTGGTGAAGTGTAAATCACTGGCTAATGAGA 2537
DB 134 ATAGCTTAGTGGTGAAGCAAGAGAGAGAGTGGTGAAGTGTAAATCACTGGCTAATGAGA 193
QY 2538 AAAGGAGACAGCTAACTCTTAGGATGAAGCTGTGACTAGGCTGGAGTTGCTTCTTGAAGA 2597

DB 194 AAAGGAGACAGCTAACTCTTAGGATGAAGCTGTGACTAGGCTGGAGTTGCTTCTTGAAGA 253
QY 2598 TGGGACTCTCTGGGTATCAAGACCTATGCCACATCATCTGGGGCTAGGAGTAGGTGA 2657
DB 254 TGGGACTCTCTGGGTATCAAGACCTATGCCACATCATCTGGGGCTAGGAGTAGGTGA 313
QY 2658 TCCAGCCCTCAAGTCTGTCTTTCAGCCAGGAGCTTGAGAGTTATATTGGGAGTGGCTC 2717
DB 314 TCCAGCCCTCAAGTCTGTCTTTCAGCCAGGAGCTTGAGAGTTATATTGGGAGTGGCTC 373
QY 2718 CAATCTGTGGACCATGATTTTCAGCTTTCCTGAAGATCAGGAGGTGCCATTCATGTC 2777
DB 374 CAATCTGTGGACCATGATTTTCAGCTTTCCTGAAGATCAGGAGGTGCCATTCATGTC 433
QY 2778 TTTCTCTCTAGCCCTCAGGAAAGAGAGTATATTGTACTGTGTACCTAGGGGTCTT 2837
DB 434 TTTCTCTCTAGCCCTCAGGAAAGAGAGTATATTGTACTGTGTACCTAGGGGTCTT 493
QY 2838 GGAAGGAAACATGGAATCTATAGACTGATAGGCCCTATCCACAGGSCCAT 2897
DB 494 GGAAGGAAACATGGAATCTATAGACTGATAGGCCCTATCCACAGGSCCAT 553
QY 2898 GACTGGGAAAGGTATGGGAGCAGAGAGAAATTTGGGATTTAGGTGAGCTACCTCA 2957
DB 554 GACTGGGAAAGGTATGGGAGCAGAGAGAAATTTGGGATTTAGGTGAGCTACCTCA 612
QY 2958 CCTTAACTTTTGGTGGCTGGGCGATGCTTGAGGCCAGAGCTGTTAAGCAGGCTCTGC 3017
DB 613 CCTTAACTTTTGGTGGCTGGGCGATGCTTGAGGCCAGAGCTGTTAAGCAGGCTCTGC 672
QY 3018 TGGCTGTCTTACTGCTCACCACCTCTGCACCTGCTCTGTGAGACTTCCAGCCCCAG 3077
DB 673 TGGCTGTCTTACTGCTCACCACCTCTGCACCTGCTCTGTGAGACTTCCAGCCCCAG 732
QY 3078 GCAGCCACCTGCTCTGAGCCCTCCTACTATCTCCCTGTGACGGTGAATCTCGTGTACTG 3137
DB 733 GCAGCCACCTGCTCTGAGCCCTCCTACTATCTCCCTGTGACGGTGAATCTCGTGTACTG 792
QY 3138 TGTCTCGGTCATATATGAATTTGAGCAGGCTCATCTATTATTAACACAGATCTTTA 3197
DB 793 TGTCTCGGTCATATATGAATTTGAGCAGGCTCATCTATTATTAACACAGATCTTTA 852
QY 3198 CAAATAAAGATTATTTCAAAACACCAAAAAA 3240
DB 853 CAAATAAAGATTATTTCAAAACACCAAAAAA 895
RESULT 9
AAZ27472
ID AAZ27472 standard; cDNA; 566 BP.
XX
AC AAZ27472;
XX
DT 07-NOV-2001 (first entry)
XX
DE cDNA encoding novel signal transduction pathway protein, Seq ID 507.
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;
XX acquired immune deficiency syndrome.
OS Homo sapiens.
XX
PN WO200154733-A1.
XX

PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01312.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-02114886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0215647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226581.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
XX
PI (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465460/50.
DR

DR P-PSDB; AAU17555.
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders
XX
XX
XX Claim 1; SEQ ID No 507; 880pp; English.
PS
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
CC pathway protein coding sequences and PCR primers of the invention.
XX

Query Match 16.4%; Score 534.2; DB 22; Length 566;
Best Local Similarity 98.9%; Pred. No. 1.3e-119;
Matches 533; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 GCCCGCTCCGCCGCCGAGTTCCTCCGCCGCCGCTGGCCCGCCAGTCATGGCGAAGCAGTACGA 61
Db 28 GCCCGCTCCGCCGCCGAGTTCCTCCGCCGCCGCTGGCCCGCCAGTCATGGCGAAGCAGTACGA 87
QY 62 TGTGCTGTTCGGCTGCTGTCGATCGGGACTCCGGGTGGCGACACCTGCTGCTGTG 121
Db 88 TGTGCTGTTCGGCTGCTGTCGATCGGGACTCCGGGTGGCGACACCTGCTGCTGTG 147
QY 122 CGCCTTCACCGACACGAGTTCCTCTCGCACATCTCCACATCGGTGTGTGACTTAA 181
Db 148 CGCCTTCACCGACACGAGTTCCTCTCGCACATCTCCACATCGGTGTGTGACTTAA 207
QY 182 GATGAAGACCATAGAGGTAGCGGCATCAAAGTGGCGATACAGATCTGGGACATGCGAGG 241
Db 208 GATGAAGACCATAGAGGTAGCGGCATCAAAGTGGCGATACAGATCTGGGACATGCGAGG 267
QY 242 GCAGGAGATACAGACCATCACAAAGCAGTACTATCGGGGGCCCGAGGATATTTT 301
Db 268 GCAGGAGATACAGACCATCACAAAGCAGTACTATCGGGGGCCCGAGGATATTTT 327
QY 302 GGTCTATGACATTAGCAGCGCGCTCTTACCAGCACATCATGAAGTGGTTCAGTACGT 361
Db 328 GGTCTATGACATTAGCAGCGCGCTCTTACCAGCACATCATGAAGTGGTTCAGTACGT 387
QY 362 GGATGAGTACGCACAGAGCGCTCCAGAAATCTCTATTGGGAATAAGGCTGATGAGGA 421
Db 388 GGATGAGTACGCACAGAGCGCTCCAGAAATCTCTATTGGGAATAAGGCTGATGAGGA 447
QY 422 GCAGAAACGGCAGGTGGGAAGAGCAAGGGCAGCGCTGGCGAAGAGATATGGCATGGA 481
Db 448 GCAGAAACGGCAGGTGGGAAGAGCAAGGGCAGCGCTGGCGAAGAGATATGGCATGGA 507
QY 482 CTTCTATGAACAAGTGGCTGCACCAACTCAACATTAAGAGATCATTCACCGCTCTGA 540
Db 508 CTTCTATGAACAAGTGGCTGCACCAACTCAACATTAAGAGATCATTCACCGCTCTGA 566

RESULT 10
ABK44752
ID ABK44752 standard; cDNA; 481 BP.
XX
XX AC ABK44752;
XX
XX DT 05-JUN-2002 (first entry)
XX
XX CDNA encoding colon tumour protein, SEQ ID No 303.
XX
XX Human; colon tumour; vaccine; colon cancer; immunogenic;
KW immunotherapy; gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200212328-A2.
XX
XX PD 14-FEB-2002.
XX
XX PF 31-JUL-2001; 2001WO-US24218.
XX
XX PR 03-AUG-2000; 2000US-223283P.
XX
XX PR 28-MAR-2001; 2001US-279763P.
XX
XX PR 29-JUN-2001; 2001US-302051P.
XX
XX PA (CORI-) CORIXA CORP.
XX
XX PI King GE, Meagher MJ, Xu J, Secríst H;
XX
XX WPI; 2002-241739/29.
XX
XX New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT for diagnosing, preventing, and treating colon cancer, and as markers
PT for the progression of cancer
XX
XX Claim 1; SEQ ID No 303; 147pp; English.
XX
XX The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABK44450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.
CC Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC for this patent did not form part of the printed specification but was
CC supplied by the European Patent Office.
XX
XX Sequence 481 BP; 111 A; 113 C; 130 G; 125 T; 2 other;

Query Match 14.7%; Score 477.4; DB 24; Length 481;
Best Local Similarity 99.4%; Pred. No. 7.3e-106;
Matches 478; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2582 GTTGCTTCCTTGAAGATGGGACTCCTTGGGTATCAAGACCTATGCACATCACACTGGG 2641
Db 1 GTTGCTTCCTTGAAGATGGGACTCCTTGGGTATCAAGACCTATGCACATCACACTGGG 60
QY 2642 CTAGGGAAGTAGTGTATGCCAGCCCTCAAGTCTGCTTCAGCCAGGACTTCAGAGTTA 2701
Db 61 CTAGGGAAGTAGTGTATGCCAGCCCTCAAGTCTGCTTCAGCCAGGACTTCAGAGTTA 120
QY 2702 TATTGGGAGTGGCTCCCAATCTGTGGACACAGTATTTTCAGCTTTCCTGAAGATCAGGCAG 2761
Db 121 TATTGGGAGTGGCTCCCAATCTGTGGACACAGTATTTTCAGCTTTCCTGAAGATCAGGCAG 180
QY 2762 GGTGCAATTCATGTCTTTCTCTCTAGCCCTCAGGAAAGAGACTATATTTGTACT 2821
Db 181 GGTGCAATTCATGTCTTTCTCTCTAGCCCTCAGGAAAGAGACTATATTTGTACT 240

Db 154 CTCCTGAGCCTCCACTATCTCCCTGTGAGCGGTGAACCTCGTACTGTCTCGGTCC 95
Qy 3150 ATATATGAATGTGAGCAGGTTCTATCTATTTAAACACACAGATGTTTACAAAATAAAGAT 3209
Db 94 ATATATGAATGTGAGCAGGTTCTATCTATTTAAACACACAGATGTTTACAAAATAAAGAT 35
Qy 3210 TATTTCAACACCACCAAAAAAATAAAAAA 3243
Db 34 TATTTCAACACCACCAAAAAAATAAAAAA 1
RESULT 12
ID AAT25457 standard; cDNA to mRNA; 320 BP.
XX AC AAT25457;
XX DT 17-OCT-1996 (first entry)
DE Human gene signature HUMGS07624.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX OS Homo sapiens.
XX PN W09514772-A1.
XX PD 01-JUN-1995.
XX PF 11-NOV-1994; 94WO-JP01916.
PR 12-NOV-1993; 93JP-0355504.
XX PA (MATS/) MATSUBARA K.
XX PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
XX WPI; 1995-206931/27.
DR Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX Claim 1; Page 1835; 2245pp; Japanese.
XX A single-stranded DNA (or its complementary strand or the corresp.
XX double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAT19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 320 BP; 74 A; 75 C; 86 G; 81 T; 4 other;
Query Match 9.6%; Score 313.4; DB 16; Length 320;
Best Local Similarity 98.7%; Pred. No. 4.1e-66;
Matches 314; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 2752 GATCAGCAGGTTGCCAATTCATTTCTCTCTAGCCCCCTCAGGAAAGAGGACTA 2811

Db 1 GATCAGCAGGTTGCCAATTCATTTCTCTCTAGCCCCCTCAGGAAAGAGGACTA 60
Qy 2812 TATTTGTACTGTACCTAGGGTTCTCGAAGGAAAACATGAATCAGGATTCATAGAC 2871
Db 61 TATTTGTACTGTACCTAGGGTTCTCGAAGGAAAACATGAATCAGGATTCATAGAC 120
Qy 2872 TCATAGGCCCTATCCACAAGGCCCATGACTGGAAAAAGGTATGGGAGCAGAGGAGAAT 2931
Db 121 TCATAGGCCCTATCCACAAGGCCCATGACTGGAAAAAGGTATGGGAGCAGAGGAGAAT 180
Qy 2932 GGGATTTTAGGTGCAGCTACGCTCACCTAACTTTTGGTGGCCTGGGGCATGTCTGA 2991
Db 181 GGGATTTTAGGTGCAGCTACGCTCACCTAACTTTTGGTGGCCTGGGGCATGTCTGA 240
Qy 2992 GGGCCAGACTGTTAAGCAGGCTCTGCTGGCCTGTTTACTCGTCACACCTCTGCACCTGC 3051
Db 241 GGGCCAGACTGTTAAGCAGGCTCTGCTGGCCTGTTTACTCGTCACACCTCTGCACCTGC 300
Qy 3052 TGTCTTGAGACTCCATCC 3069
Db 301 TGTCTTGAGACTCCATCC 318
RESULT 13
ID ABL38341 standard; cDNA; 313 BP.
XX AC ABL38341;
XX DT 08-APR-2002 (first entry)
DE Human colon tumour antigen polynucleotide SEQ ID NO:1930.
XX KW Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
KW colon tumour metastatic antigen; diagnosis; gene; ss.
XX OS Homo sapiens.
XX PN W0200196388-A2.
XX PD 20-DEC-2001.
XX PF 08-JUN-2001; 2001WO-US18557.
XX PR 09-JUN-2000; 2000US-210899P.
PR 20-FEB-2001; 2001US-270216P.
XX PA (CORI-) CORIXA CORP.
XX PI Jiang Y, Harlocker SL, Secrlist H;
XX WPI; 2002-114514/15.
DR Novel isolated colon tumor polynucleotide differentially expressed in
PT colon tumor or colon metastatic tumor and polypeptides encoded by them,
PT useful for inhibiting development of cancer in patient -
XX Claim 1; SEQ ID 1930; 105pp; English.
XX ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I)
CC which were isolated from human colon tumour and colon metastatic tumour
CC cDNA libraries. (I) have cytostatic activity and can be used in vaccine
CC production. (I) can be used for stimulating and/or expanding T cells
CC specific for a tumour protein on contact with the T cells. They are also
CC useful for inhibiting the development of cancer in a patient. (I) can be
CC used as probes or primers for nucleic acid hybridisation, for preparing
CC mutant species primers, or primers for use in genetic constructions. (I)
CC can be used in the diagnosis of a colon tumour.
SQ Sequence 313 BP; 68 A; 85 C; 85 G; 73 T; 2 other;
Query Match 9.4%; Score 307; DB 24; Length 313;

Best Local Similarity 99.0%; Pred. No. 1.4e-64;		Matches 308; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	
QY	2824	ACCTAGGGGTTCTGGAAGGAAACATGAATCAGGATTCATAGACTAGAGCCCTA	2883
Db	1	ACCTAGGGGTTCTGGAAGGAAACATGAATCAGGATTCATAGACTAGAGCCCTA	60
QY	2884	TCACAAGGCCATGACTGGGAAAGGTATGGAGCAGAGAAATGGGATTTAGGG	2943
Db	61	TCACAAGGCCATGACTGGGAAAGGTATGGAGCAGAGAAATGGGATTTAGGG	120
QY	2944	TGCAGTAGCTCACCTAACTTTGGTGGCTGGGCGATGCTTGAGGCCAGACTGT	3003
Db	121	TGCAGTAGCTCACCTAACTTTGGTGGCTGGGCGATGCTTGAGGCCAGACTGT	180
QY	3004	TAGCAGGCTCTGCTGGCTGTTTACTGCTGCACCACTCTGCACCTGCTTGGAGACT	3063
Db	181	TAGCAGGCTCTGCTGGCTGTTTACTGCTGCACCACTCTGCACCTGCTTGGAGACT	240
QY	3064	CCATCAGGCCAGGACGCCACCTGCTGAGCCTCCACTATCTCCCTGTGACGGTG	3123
Db	241	CCATCAGGCCAGGACGCCACCTGCTGAGCCTCCACTATCTCCCTGTGACGGTG	300
QY	3124	AACTTCGTGTA 3134	
Db	301	AACTTCGTGTD 311	
RESULT 14			
ABA09160			
ID	ABA09160 standard; cDNA; 1986 BP.		
XX	ABA09160;		
XX	11-JAN-2002 (first entry)		
DE	Human rab8 homologue-encoding cDNA, SEQ ID NO:936.		
XX	Human; cytokine; cell proliferation; cell differentiation; growth factor;		
XX	haematopoiesis regulation; tissue growth; immunomodulator; activin;		
KW	inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;		
KW	proliferation; metastasis; cancer; tumour; haematopoietic disorder;		
KW	myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;		
KW	chronic inflammatory condition; proliferative retinopathy;		
KW	atherosclerosis; coronary heart disease; arterial ischaemia;		
KW	bone disorder; osteoporosis; vascular growth disorder;		
KW	tissue regeneration; wound healing; infection; immune disorder;		
KW	cell culture; drug screening; gene therapy; antiinflammatory;		
KW	antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;		
KW	cytostatic; osteoparalic; vasotropic; cardiant; virucide; antibacterial;		
KW	antifungal; vulnery; antiulcer; ss.		
XX	Homo sapiens.		
OS	WO200157188-A2.		
XX	09-AUG-2001.		
XX	05-FEB-2001; 2001WO-0503800.		
XX	03-FEB-2000; 2000US-0496914.		
PR	27-APR-2000; 2000US-0560875.		
XX	(HYSE-) HYSEQ INC.		
PA	Tang YT, Liu C, Drmanac RT;		
XX	WPI; 2001-457740/49.		
DR	P-PSDB; ABB11916.		
XX	Human proteins and DNA encoding sequences useful for preventing,		
PT	treating or ameliorating a medical condition in a mammalian subject		
PT	e.g. arthritis and cancer		

XX	Claim 1; Page 805; 1963pp; English.		
XX	Sequences ABB10981-ABBI2330 represent 1350 novel human polypeptides, and sequences ABA08225-ABA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoiesis regulatory activity; tissue growth factor activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a cDNA encoding a novel human polypeptide of the invention.		
XX	Sequence 1986 BP; 519 A; 545 C; 469 G; 453 T; 0 other;		
SQ	Query Match 7.0%; Score 229; DB 22; Length 1986;		
	Best Local Similarity 67.0%; Pred. No. 2.8e-45;		
	Matches 325; Conservative 0; Mismatches 160; Indels 0; Gaps 0;		
QY	45	ATGGCGAAGCAGTACGATGTCTGTCGGGCTGCTGTCGTCGGGACTCCGGGTGGGC	104
Db	19	ATGGCGAAGCAGTACGATGTCTGTCGGGCTGCTGTCGTCGGGACTCCGGGTGGGC	78
QY	105	AAGACCTGCTGCTGTGCGGCTTCCACGACACAGAGTTCACCTCCCGACATCTCCACC	164
Db	79	AAGACCTGCTGCTGTGCGGCTTCCCGAGGAGCGCTTCAACTCCACTTTTATCTCCACC	138
QY	165	ATCGGTGTTGACCTTTAAGATGAAGACCATAGAGGTAGACGGCATCAAGTCGGATACAG	224
Db	139	ATAGGAATTGACCTTTAAATATTAGGACCATAGAGCTCGATGGCAAGAGATAAATACATGCAG	198
QY	225	ATCTGGGACACTGCAGGGCAGAGAGATACACGACCATCACAAAGCAGTACTATCGCGCG	284
Db	199	ATATGGGACACACCGGTCAGGAACGGTTTCGGACGATCACACGCGCTACTACAGGGGT	258
QY	285	GCCCCAGGGGATATTTTGGTCTATGACATTTAGCAGGCGGCGCTCTTACCAGCACATCATG	344
Db	259	GCAATGGGCATCATGCTGTGTTCTACGACATCACCACGAGAGAGTCTTCGACACATCCGG	318
QY	345	AAGTGGTCACTGACGTGGATGAGTAGCGCACAGAGCGCTCCAGAGATCCCTATTATGGG	404
Db	319	AACTGGATTGCAACATTTAGAGGAGCAGCGCTCTGACAGCTCGAAAAGATGATACTCGGG	378
QY	405	AATAAGCTGATGAGGAGCAAAACGCGAGGTGGGAGAGAGAGCAAGGCGACGAGTGGCG	464

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2002, 16:48:24 ; Search time 5516 Seconds
(without alignments)
17184.162 Million cell updates/sec

Title: US-09-817-198A-1

Perfect score: 3257

Sequence: 1 tgccgcgtgccgccgcag.....aaaaaaaaaaaaaaaaaaaa 3257

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description	
No.	%							
1	2696.6	82.8	190517	9	CNS01DX4	AL139022	Human chr	
2	1192.8	36.6	7924	6	AX347040	AX347040	Sequence	
3	1192.8	36.6	7924	6	AX348456	AX348456	Sequence	
C	4	1160.8	35.6	7924	6	AX347041	AX347041	Sequence
C	5	1160.8	35.6	7924	6	AX348457	AX348457	Sequence
6	1000.2	30.7	1054	6	AX399903	AX399903	Sequence	
7	866.2	26.6	895	6	AX014147	AX014147	Sequence	
8	813	25.0	3139	10	BC027769	BC027769	Mus muscu	
9	577.6	17.7	945	10	RATRA15X	M83679	Sprague-Daw	
10	477.4	14.7	481	6	AX396088	AX396088	Sequence	
C	11	438.8	13.5	463	6	AX334820	AX334820	Sequence
C	12	423.6	13.0	162495	2	AC096084	AC096084	Rattus no
13	416.4	12.8	2560	10	BC013790	BC013790	Mus muscu	
C	14	378.4	11.6	257226	2	AC124556	AC124556	Mus muscu
15	307	9.4	313	6	AX341683	AX341683	Sequence	
16	234.2	7.2	2048	9	BC002977	BC002977	Homo sapi	
17	234.2	7.2	2818	9	AK025165	AK025165	Homo sapi	
18	230.6	7.1	1980	9	S53268	S53268	Homo sapien	
19	229	7.0	624	9	AF498943	AF498943	Homo sapi	
20	229	7.0	660	9	HSRAB8	HSRAB8	H.sapiens m	
21	227.4	7.0	638	10	S53270	S53270	MEL=RAS-rel	
22	224.2	6.9	1337	10	BC019990	BC019990	Mus muscu	
23	222.6	6.8	760	4	CFRAB8	X56385	Canine rab8	
24	218.2	6.7	765	5	DYGORA2	M38391	Discopyge o	
25	217.8	6.7	624	6	AX236078	AX236078	Sequence	
26	217.8	6.7	1128	9	BC020654	BC020654	Homo sapi	
27	217.8	6.7	1161	6	AX236076	AX236076	Sequence	
28	217.8	6.7	1265	9	AB038995	AB038995	Homo sapi	
29	217.8	6.7	2497	6	AX285074	AX285074	Sequence	
30	217.8	6.7	2497	6	AX285080	AX285080	Sequence	
31	217.8	6.7	2497	6	AX285089	AX285089	Sequence	
32	217.8	6.7	2497	6	AX285091	AX285091	Sequence	
33	217.8	6.7	3077	9	AK001111	AK001111	Homo sapi	
34	216.2	6.6	3110	9	HSB04678	AL833365	Homo sapi	
35	211.4	6.5	740	10	RNU53475	U53475	Rattus norv	
36	203.2	6.2	991	10	RATRA10X	M83677	Sprague-Daw	
37	201.8	6.2	897	10	AF035646	AF035646	Mus muscu	
C	38	198.6	6.1	110000	2	AC048347_0	AC048347	Homo sapi
C	39	198.6	6.1	110000	2	AC048347_1	Continuation (2 of	
40	198.6	6.1	302718	2	AC027292	AC027292	Homo sapi	
41	198.4	6.1	1029	9	BC000896	BC000896	Homo sapi	
42	198.4	6.1	3533	9	AK023223	AK023223	Homo sapi	
43	196.8	6.0	3521	9	HSB01620	AL136650	Homo sapi	
44	196.6	6.0	2048	3	AY069671	AY069671	Drosophila	
45	196.6	6.0	2210	3	D84347	D84347	Drosophila	

ALIGNMENTS

RESULT 1
CNS01DX4
LOCUS
DEFINITION Human chromosome 14 DNA sequence BAC R-840119 of library RPCI-11
ACCESSION ALI39022
VERSION HTG.
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 190517)
Hellig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
Brottier, P., Cattolico, L., Barbe, V., Pelletier, E., Attienave, F.,
190517 bp DNA linear PRI 17-MAY-2001
from chromosome 14 of Homo sapiens (Human), complete sequence.

||||| GTCTTGAGACTCCATCCAGCCCGAGCAGCCACCTGCTCTCGAGCTCCACTATCTCCC 160258
Db 160199
QY 3113 TGTGAGGGTGAACCTCGTGTACTGTCTCGGGTCCATATATGAATGTGAGCAGGGTT 3172
Db 160259 TGTGAGGGTGAACCTCGTGTACTGTCTCGGGTCCATATATGAATGTGAGCAGGGTT 160318
QY 3173 CATCTATTTTAAACACAGATGTTTACAAAATAAAGATTATTTCAAACACC 3223
Db 160319 CATCTATTTTAAACACAGATGTTTACAAAATAAAGATTATTTCAAACACC 160369

RESULT 2
LOCUS AX347040 7924 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 2111 from Patent WO0200928.
ACCESSION AX347040
VERSION AX347040.1 GI:18494928
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.

REFERENCE
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 2111 03-JAN-2002;
Epigenomics AG (DE)

FEATURES
source Location/Qualifiers
1..7924 /organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2087 a 83 c 2037 g 3717 t
ORIGIN

Query Match 36.6%; Score 1192.8; DB 6; Length 7924;
Best Local Similarity 75.8%; Pred. No. 2.2e-296;
Matches 1476; Conservative 0; Mismatches 472; Indels 0; Gaps 0;

QY 1276 GGTGCTCTCTCCCTTCCCCACCCACGTCCTCCATGTCGTCATGCGCTGCGCTGCC 1335
Db 1 GGTGCT 60

QY 1336 AGTGACCTGGGAAAGTGGAGCATCGAGTAGGAGGAAACAGCAACCGGGAGTCTCGA 1395
Db 61 AGTGATTTGCGAAAGTGGAGTATCGAGGTAGGAGGAAACGGTAATTAGCGAGTTTCGA 120

QY 1396 GCCTGGGGTGCCTACCTCTACCCATTCGCCGACGAGCTTTGCCCTTGCTTGGCTGC 1455
Db 121 GTTTGGGGTGTCTTTATTTATTTATTTATTTTCGATTAGAGTTTGTCTTTTGTGGTGT 180

QY 1456 CCGCTCGCTCTTTGGGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1515
Db 181 TCGTTTGTCTTTTGGGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 240

QY 1516 GAGGGTGGCAGGATAAAAGTCACTCCATCTCTACCTCCCATGCGAGCATGAGCAACA 1575
Db 241 GAGGGTGGTAGGGATAAAAAGTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 300

QY 1576 ATTTCTCTCCACCTGGCTCCCAATTTAAAGATGTGACCAAGCCCTGTGGGTACTCCAG 1635
Db 301 ATTTCT 360

QY 1636 GGGCAAGGAGAGCCCTGGGGTCACTGACACTGTCCAGGCAACCATGCACCTCCCAAAAGG 1695
Db 361 GGGTAAGGAGAGTTTGGGGTACTGATATTTAGGTAAATATTTATTTATTTATTTATTTAG 420

QY 1696 GAGCATTTGGAATGAAGGACTAGCTCCTATGATCAGGTTAAGCAAGGAGGAGCTGG 1755
Db 421 GAGTATTTGGAATGAAGGACTAGCTCCTATGATCAGGTTAAGCAAGGAGGAGCTGG 480

QY 1756 CCAGGACAGCATTTGCACAGCAGAGGGGAATGTACCAACAGCAGGCGCTCTCCTAGGCC 1815
Db 1756 CCAGGACAGCATTTGCACAGCAGAGGGGAATGTACCAACAGCAGGCGCTCTCCTAGGCC 1815

Db 481 TTAGGATAGTAGTTGTGTATAGTAGAGGGGAATGTAGTAATAGTAGAGGTTTTTTAGGTTT 540
QY 1816 CATCTTCCATTTCTTAGGTAAAGAGACATTTCTCCTCAGACTCCCGAGCGGAGACTGAGC 1875
Db 541 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 600
QY 1876 CTAGCCTTTCAGCAACCAAGTTCTCTCTGGGACCCAAAGTTTATGGGAGAGGCGCAAGAC 1935
Db 601 TTAGTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 660
QY 1936 TTATCGGGAAG 1995
Db 661 TTTATGGGAAG 720
QY 1996 TTTAAGACAAAGCGCTCATCTTGGCTCTACCTCTACCTCTAGAGCTTGGAGGTTTGGCAACC 2055
Db 721 TTTAAGATAAAGCGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 780
QY 2056 ACACCTGTGCTACAGGTGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2115
Db 781 ATATCTGCTTATAGGTGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 2116 GTTCTCTTAACCCCATATGSCCAAGAGTAGCTCGTAGGAGGCCCTTTAAAGACGGAACA 2175
Db 841 GTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 900
QY 2176 AGTAATTTTACCAGTTCTACTGGGGTTCTCTGCCACCGCTCCCAAGGTGGCGAGGCCCTAG 2235
Db 901 AGTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 960
QY 2236 AAGAGGCTCATCTTAAAGCCACACATAGCTGCACCTGCGTGGCTGCAGCCAAACAAAGA 2295
Db 961 AAGAGGCTTATTTTAAAGTATATATAGTTTATTTTATTTTATTTTATTTTATTTTATTTT 1020
QY 2296 ACTGGGTGTGAGTATTCACTAAAGAACCAAAATTCAGGCGCTCATATCTGAAGGA 2355
Db 1021 ATTTGGGTGTGAGTATTATTAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAA 1080
QY 2356 TAAGAACTCCTCCT 2415
Db 1081 TAAGAAATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1140
QY 2416 CTCTGACTTACCAACCCAGGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2475
Db 1141 TTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1200
QY 2476 GTATAGCTTAGTGTGGGAAAGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2535
Db 1201 GTATAGTTTAGTGTGGGAAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1360
QY 2536 GAAAG 2595
Db 1261 GAAAG 1320
QY 2596 GATGGGACTCTTGGGTATCAAGACCTATGCCACATCACACTGGGCTAGGAGAGAGAGAG 2655
Db 1321 GATGGGAGTTTTTGGGTATTAAGATTATGTTATATATATGCGGTAGGAGAGAGAGAGAG 1380
QY 2656 GATGCCAGCCCTCAAGTCTCTCTTCAGCCAGGACTTTAGAAAGTTATATTTGGCAGTGGC 2715
Db 1381 GATGTTAGTTTTTAAAGTTCTCTTTAGTTAGGAGTTTGAAGAGTTATATTTGGGTAGTGT 1440
QY 2716 TCCAACTGTGGACAGATTTTCCAGCTTCCCTCGAAGATCAGGAGGAGGAGGAGGAGGAG 2775
Db 1441 TTTAAATTTGGGATAGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1500
QY 2776 TCTTCT 2835
Db 1501 TTTTCT 1560
QY 2836 CTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2895
Db 1561 TTGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620

Qy	2478	ATAGCTTAGTGTGGAAGACAAAGAGAGCAGGAAGTTGTAAATCAGCTGGCTAATGAGA	2537
Db	6722	ATAACTTAAATATAAAAAACAACAAAAATATAAATCAGTAACATAATAAA	6663
Qy	2538	AAAGGAGACAGCTAAGCTAGGATGAGAGCTGTGACTAGGCTGGAGTTGCTTCCTTGAAGA	2597
Db	6662	AAAAAACAACCTAAGCTTAAATAAAAATAAATTAACATAAACAATAAATTACTTCCCTAAAAA	6603
Qy	2598	TGGGACTCCCTGGGTATCAAGACCTATGCCACATCACACTGGGCTAGGGAAGTAGGTGA	2657
Db	6602	TAAACTCCCTTAAATATCAAAACCTATACCACTCAGCTAAAACTAAAAAATAAATAA	6543
Qy	2658	TGCCAGCCCTCAAGTCTGCTTTCAGCCAGGAGCTTGAAAGTTATATTGGGCAGTGGCTC	2717
Db	6542	TACCACCCCTCAATCTATCTTCAACCAAAAACCTTAAAAATTTATATTAAACAATACTC	6483
Qy	2718	CAATCTGTGGACAGTATTTTCAGCTTTCCTTGAAGATCAGCAGGGTGCCATTCATTTGTC	2777
Db	6482	CAATCTATAAACCAATATTTCAACTTTCCTTAAAAATCAAAACAAAATACCATTCATTATC	6423
Qy	2778	TTTCTCTCTAGCCCCCTCAGAAAGAGGAGCTATATTGTACTGTACCTTAGGGGTTCT	2837
Db	6422	TTTCTCTCTTAAAGCCCCCTCAAAAAAATAAATAATATTATTACTATACCTTAAAAATTTCT	6363
Qy	2838	GGAAGGGAACAACTGGAATTCAGGATTTCTATAGACTGATAGGCCCTATCCACAAGGGCAT	2897
Db	6362	AAAAAANAACATAAATAACAAATTTCTATAAATCTATAAACCTATAACCAATCCCAAAAACCAT	6303
Qy	2898	GACTGGGAAAAAGGTATGGGAGCAGAAGAGAAATTTGGGATTTTAGGGTGCACTACGCTCA	2957
Db	6302	AACATAAAAAAATAATAAAACAAAAAATAAATTTAAATATACACTACGCTCA	6243
Qy	2958	CCCTAAACTTTTGGTGGCCCTGGGGAATGCTTTGAGGCCAGACTGTTTAAAGCAGGCTCGC	3017
Db	6242	CCCTAAACTTTTAAATAAACCTATAAACAATATCTTTAAACCCCAAACTATTAAACCAAACTCTAC	6183
Qy	3018	TGGCTGCTTTTACCTGTCACCACTCTGCACCTGCTGCTTGAGACTCCATCCAGCCCCAG	3077
Db	6182	TAACTTATTTTACCTGTCACCACTCTACCACTTACTTCTTAAAACTCCATCCCAACCCCAA	6123
Qy	3078	GCAGGCCACTGCTCCTGAGCCTCCACTATCTCCCTGTGAGGGGTGAACCTTCGTGTAAGT	3137
Db	6122	ACAGGCCACTTCTCTAACTCCACTATCTCCCTATTAACGAATAAATCTCGTATACTA	6063
Qy	3138	TGCTCGGGTCCATATATGAATTTGTGACGAGGGTTTCATCTATTTTAAACACAGATGTTTA	3197
Db	6062	TATCTCGAATCCCATATAATAATATAAACAATAATCTATTTTAAACACAAATATTTA	6003
Qy	3198	CAAAATAAGATTTATTTCAACACCACCAAAAAAATAAATAAATAAATAAATAAATAA	3257
Db	6002	CAAAATAAATAATTTTCAACACCACCGATATAACTTACCTTAATTAATTCCTTTAAAAATATAA	5943

[illegible]

BASE COUNT	1973 a	83 c	1827 g	4041 t	Query Match	35 68;	Score 1160.8;	DB 6;	Length 7924;			
ORIGIN					Best Local Similarity	74.18;	Pred. No. 4.le-288;					
					Matches 1468;	Conservative	0;	Mismatches 512;	Indels	0;	Gaps	0;
QY	1278	TGCTTCTTCTCTTCC	CCCCACCCACCTGT	CCCTCATGTG	CCATGGCGCT	GCCTGCCCTCC	CCAG	1337				
DB	7922	TACTTCTTCTCTTCC	CCCCACCCACCTAT	CCCTCATATAC	CATTAACCTAC	CTCCCAA	7863					
QY	1338	TGACCTGCGAAGTGG	AGCATCGAGGTAG	GAGGAAACAG	CAACCGGGA	GTCTCGAGC	1397					
DB	7862	TAACTACGAAATAA	AAACATCGAAAT	AAAAA	AAACGACAA	AAAAATCCTCGAAC	7803					
QY	1398	CTGGGGCTGCCCTAC	CTTACCCATTCCC	CACAGAGCTTT	GGCCTTGC	TTGGCTGCC	1457					
DB	7802	CTAAACATACCCCTA	CTCTACCCATTCCC	ACCAAACTTT	TACCCCTTAC	TTTAATACCT	7743					
QY	1458	GCCTGCCCTCTTTT	GGGGAATCGAGCT	CAGCGAGTGT	CCAGAGGA	AAAAAATGA	1517					
DB	7742	GCCTACCTCTTTA	AAAAAACTAACT	CAAAAACAAAT	TACTTCA	AAAAA	7683					
QY	1518	GGGTGGCAGGGATA	AAAAAGTCACCT	CCATTTCTCTAC	CTCCATCG	ACATGAACAAAT	1577					
DB	7682	AAAAATAACAAAAA	TAAAAAATCAGCT	CCATTTCTAC	CTCCATACA	ATAAACACAAAT	7623					
QY	1578	TTCTCTCACCTGG	CTCCCAATTTAA	GATGTGGACCA	AGCGCTGT	GGGTACTCCAGGG	1637					
DB	7622	TTCTCTCACCTTA	CTCCCAATTTAA	AAATATAAAC	CAAAACCTT	ATAATACTCCAAA	7563					
QY	1638	GCAAGGACAGCC	TGGGGTCACTG	CAGTCACTG	CAGSCCA	ACCATGCAC	1697					
DB	7562	ACAAAAA	AAACCCTA	AAATCAATAC	ACTATCA	AAACCAACCATCAC	7503					
QY	1698	GCATTTGGAAATG	AAGGACTAGCT	CCTCTATGAT	CAGGTTA	AGACGAAGGAGCTGCC	1757					
DB	7502	ACATTTAAAAAT	AAAAAACTAACT	CTCTATATCA	AAATTTAA	AAAAA	7443					
QY	1758	AGGNACACGATTT	GCACAGCAGGG	GAATGTAG	CAACACAG	CGGCCCTCTAGGCCCA	1817					
DB	7442	AAAAACAACAAT	TACACACAAAA	AAAAATATAC	ACAACA	AAACCTCTTAACCCCA	7383					
QY	1818	TCTTCCATTTCTT	AGTAAAGAGAC	ATTCTCTCAG	ACTCC	CAGCGGAGACTGAGCT	1877					
DB	7382	TCTTCCATTTCTT	AAATAAAAA	AAACATTTCT	CNAAC	TCCCAACGAAAACTTAACT	7323					
QY	1878	AGCCTTCAGCAAC	CAAGGTTCTCT	CTGGACCCAA	AGTTTAT	GGGAGAGGCAAGACTT	1937					
DB	7322	AACCTTCAACAC	CAAAATTTCT	CTCTAAAC	CCAAAAATTT	TAAAAA	7263					
QY	1938	CATGGGAAGAG	AGAAAGGCCCTGG	TAGAAACGCTTGG	TGCTGTCT	CTTTGGCCTT	1997					
DB	7262	CATAAAAA	AAAAA	AAAAACCC	TAAATA	AAAAAGCTTAAT	7203					
QY	1998	TAAACAAGCGCTCA	CTCTTGCCTCTAC	CTCTCTGAT	AGGCTTG	GAGGTTTGC	2057					
DB	7202	TAAACA	AAACGCTCATCTT	ACCTCTAC	CTCTTA	ATAAATTA	7143					
QY	2058	ACTGTGGCTAC	AGGTGGGAAGG	AGGACTCCTT	CTCCAGAGT	GCTATGTCAGGAAGT	2117					
DB	7142	ACTATACTACA	ATAAAAA	AAAAAA	AACTCCTT	CCAAAAAT	7083					
QY	2118	TTCTTTAACCC	ATATGCCCCA	AGATGAGT	CTAGGAGCC	CTTTAAAGACGGAACAG	2177					
DB	7082	TTCTTTAACCC	ATATACCC	AAAAA	TAACTCTG	TAAAAA	7023					
QY	2178	TAAATAC	CAAGTTCTACT	GGGTTCTCT	CCACCGTCC	CAAGTGGGCGAGGCTTAGAA	2237					
DB	7022	TAAATACCA	ATTTCTACT	AAAAATCTCT	TACCCACCGTCC	CAAAATTA	6963					
QY	2238	GAGGCTCATTTCT	TAGCCACACAT	TAGCTG	CACTCGCTG	CGCTGAGCGGCAACAAAGAAC	2297					

Db	6962	AAAAATCATTTCTAAACACACATTAACCTACACTACGTAACTACACCAAAAAACAAAAAC	6903
Qy	2298	TGGGTCTTCAGTATTCATCAACTAGAACCAAAATCCAGGCGACATCATATGTGAAGGATA	2357
Db	6902	TAAATATTAATATTCATCACTCACTAAACCAAAATCCAAACACATCATATATAAAAAATA	6843
Qy	2358	AGAACCTCACTTCCTTACTCTCCAAAAAGAGTGGGAAAGAACCATCAAAACCTTTCCT	2417
Db	6842	AAAACCTCACTTCCTTACTCTCCAAAAAAATAAAAAAAACCATCAAAACCTTTCCT	6783
Qy	2418	CCTGACTTACCAAAACAGGAACACGACGAGAGGGTGCTCAGGACTTAGGCAAGGCT	2477
Db	6782	CCTAACTTACCAACCAAAAAACACAAAAAAATAAATCAAACTTTAAAAACAAAAAT	6723
Qy	2478	ATAGCTTACATGGTGAAGAGAGAGAGAGAGTGTAAATCACGTGGCTAATGAGA	2537
Db	6722	ATAACTTAATATAAAAAACAAAAAAACAAAAAATATATAATCACTAACTAATAAAA	6663
Qy	2538	AAAGGACGAGCTAATCTTAGGATGAAGCTGTGACTTAGCTGGAGTTGCTTCTTTGAAGA	2597
Db	6662	AAAAAAACAACCTAATCTTCAACCAAAACCTTAAAAAATATATTAACAATAAATC	6603
Qy	2598	TGGGACTCCTTGGGTATCAAGACCTATGCCATACACTGGGCTAGGAACTAGGTGA	2657
Db	6602	TAAACTCCTTAATATCAAAACCTATACCACATCACATAAACTAAAAAATAAATAA	6543
Qy	2658	TGCCAGCCCTCAAGTCTGTCTTCAGCGAGGACTTCAGAAAGTTATATTGGCGAGTGGCT	2717
Db	6542	TACCAACCCCTCAATCTATCTTCAACCAAAACCTTAAAAAATATATTAACAATAAATC	6483
Qy	2718	CAATCTGTGACACGATTTTCAGCTTTCCCTGAAGATCAGGCGAGGTCGCATTCAATGTC	2777
Db	6482	CAATCTATAACCAATATTTCAACTTTCCCTAAAAATCAAAACAAAAATACCAATTCATTATC	6423
Qy	2778	TTTCTCTCTAGCCCTCAGGAAGAACGACTATATTTGTACTGTACCTGAGGGGTTCT	2837
Db	6422	TTTCTCTCTAAGCCCTCAAAAAAACAACATATATTTATACATACCCCTAAAAATCT	6363
Qy	2838	GGAAGGAAAAACATGAATCAGGATTTCTATAGACTGATAGGCTATCCACAAAGGCCAT	2897
Db	6362	AAAAAAACCAATAAAATCAAAATTTCTATAAATAATAAACCCTATCCACAAAAACCAT	6303
Qy	2898	GACTGGAAAGATGTGGGACGAGAGAGATTTGGGATTTTAGGTGCGAGCTACGCTCA	2957
Db	6302	AACTAAAAAATAATATAAAACAAAAAATAAATTTTAAAAATACAACTACGCTCA	6243
Qy	2958	CCCTAAACTTTTGGTGGCCTGGGGCATGTCTTGAGGCCAGACTGTTAAGCAGGCTCTGC	3017
Db	6242	CCCTAAACTTTTAACTTAAACATATCTTAAACCCCAAACTATTAAACCAACTCTAC	6183
Qy	3018	TGGCTGTTTACTCGTCACCACTCTGCACCTGCTCTTGAGACTCCATCCAGCCCCAG	3077
Db	6182	TAACCTATTACTCGTCACCACTCTACACCTACTATCTTTAAACCTCCATCCAAACCCCA	6123
Qy	3078	GCAGGCACCTGCTCCTGAGCCTCCTACTATCTCCCTGTGACGGGTGAACCTGCTGACTG	3137
Db	6122	ACAGCCACCTACTCTCTAAACCTCCACTATCTCCCTATAACGAATAAACTTCGTATACTA	6063
Qy	3138	TGTCCTGGGTCCATATATGAATTTGTGAGCAGGCTTCATCTATTTTAAACACAGATGTTA	3197
Db	6062	TATCTGAAATCCATATATAAATTTATAACAAAATTCATCTATTTTTAAACACAATATTTA	6003
Qy	3198	CAAAATAAAGATTTATTTCAACACCACCAAAAAAATAAAAAAATAAAAAAATAAAAAA	3257
Db	6002	CAAAATAAATAATTTTCAAAACCAACCGATATAACTACCTAAATAAATCCCTTAAAAATAA	5943

RESULT 6
AX399903
LOCUS
DEFINITION Sequence 74 from Patent WO0218424.
ACCESSION AX399903

AX399903 1054 bp DNA linear PAT 06-JUN-2002

VERSION	AX399903.1	GI:21336157
KEYWORDS	human.	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1	
AUTHORS	Tang, Y. T., Asundi, V., Zhou, P., Xue, A. J., Ren, F., Zhang, J., Wang, J. R., Zhao, Q. A., Wang, D., Liu, C., Drmanac, R. T. and Wehrman, T.	
TITLE	Nucleic acids and polypeptides	
JOURNAL	Patent: WO 0218424-A 74 07-MAR-2002;	
FEATURES	HYSEQ, INC. (US)	
source	Location/Qualifiers	
	1..1054	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
CDS	81..719	
	/note="unnamed protein product"	
	/codon_start=1	
	/protein_id="CAD33347.1"	
	/db_xref="GI:21336158"	
	/translation="MAKQYDLVRLLLIGDSGVKTKLLCRFTDNEFHSHSTIGVD FKMTIEVDGIKRVIOIWDTAGOERYOTITKOYVRAOGIFLVYDISSERSVOHIMKW VSDVDEYAPGVOKILIGNKADEOKROVREGOOLAKEXYMDRYETSACTNLNIKE SFRLTELVLQAHKLEBGLRMRASNELALAELEEBEGKPEGPANSKTWC"	
BASE COUNT	240 a 314 c 304 g 195 t	
ORIGIN		
	Query Match 30.7%; Score 1000.2; DB 6; Length 1054;	
	Best Local Similarity 99.6%; Pred. No. 8.2e-247;	
	Matches 1002; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
QY	2	GCCCGCTGCCCGCCCGCAGTTCCCGGCCCGCTGGCCCCCAGTCATGGCAAGCAGTACGA 61
Db	38	GCCCGCTGCCCGCCCGCAGTTCCCGGCCCGCTGGCCCCCAGTCATGGCAAGCAGTACGA 97
QY	62	TGTGCTGTTCCGGCTGCTGCTGATCGGGGACTCCGGGTGGCAAGACCTGCTGCTGTG 121
Db	98	TGTGCTGTTCCGGCTGCTGCTGATCGGGGACTCCGGGTGGCAAGACCTGCTGCTGTG 157
QY	122	CGCTTTCACCGCAACGAGTTCACCTCCCGACATCTCCACCATCGGTGTTGACTTTAA 181
Db	158	CGCTTTCACCGCAACGAGTTCACCTCCCGACATCTCCACCATCGGTGTTGACTTTAA 217
QY	182	GATGAAGACCATAGAGGTAGCGGCATCAAGTGGCGATACAGATCTGGGACACTGCAGG 241
Db	218	GATGAAGACCATAGAGGTAGCGGCATCAAGTGGCGATACAGATCTGGGACACTGCAGG 277
QY	242	GCAGGAGATACCCAGCCATCACAAAGCAGTACTATCGCGGGCCCGAGGGATATTTT 301
Db	278	GCAGGAGATACCCAGCCATCACAAAGCAGTACTATCGCGGGCCCGAGGGATATTTT 337
QY	302	GGTCTATGACATTAGCAGCGAGCGCTTTACCAGCACATCATGAAGTGGTCTAGTGACGT 361
Db	338	GGTCTATGACATTAGCAGCGAGCGCTTTACCAGCACATCATGAAGTGGTCTAGTGACGT 397
QY	362	GATGAGTAGCAGCACAGAGCGGTCAGAGATCCCTTATTTGGGAATAAGCTCATGAGGA 421
Db	398	GGATGAGTAGCAGCACAGAGCGGTCAGAGATCCCTTATTTGGGAATAAGCTCATGAGGA 457
QY	422	GCAGAAACGCGAGCTGGGAAGAGAGCAAGGCGACGAGCTGGCCGAAGAGTATGGCATGGA 481
Db	458	GCAGAAACGCGAGCTGGGAAGAGAGCAAGGCGACGAGCTGGCCGAAGAGTATGGCATGGA 517
QY	482	CTTCTATGAACAAGTGCCTGCACCAACCTCAACATTTAAAGAGTCATTCACGCGTCTGAC 541
Db	518	CTTCTATGAACAAGTGCCTGCACCAACCTCAACATTTAAAGAGTCATTCACGCGTCTGAC 577
QY	542	AGAGCTGGTGTCTGCAGGCCCCATAGGAAGAGCTGGGAAGGCTCCGGATGGTCCAGCAA 601
Db	578	AGAGCTGGTGTCTGCAGGCCCCATAGGAAGAGCTGGGAAGGCTCCGGATGGTCCAGCAA 637
QY	602	TGAGTTGGCACTGGCAGAGCTGGGAGGAGGAGGAGGCAAAACCCGAGGCGCCCGCAACTC 661


```
|||||
Db 638 TCAGTTGGCACTGGCAGAGCTGGAGGAGGAGGAGGCAAAACCCGAGGGCCCGAGAACTC 697
QY 662 TTCGAAACCTTCCTGGTCTGAGTCTCTGTGTGGGCGACCCACACAGACACCCCTTCTCC 721
Db 698 TTCGAAACCTTCCTGGTCTGAGTCTCTGTGTGGGCGACCCACACAGACACCCCTTCTCC 757
QY 722 TCAGGAGCCGCTGGGCGACAGAGGGAGCGGGGCTTTGCCCTGTCTGTCTCTCTCGTG 781
Db 758 TCAGGAGCCGCTGGGCGACAGAGGGAGCGGGGCTTTGCCCTGTCTGTCTCTCTCATG 817
QY 782 TGATGACCCCTATTTAGTATCATAGTACCTACTCTCCCTCTGCTGTCTGTCTGTCTGTG 841
Db 818 TGATGACCCCTATTTAGTATCATAGTACCTACTCTCCCTCTGCTGTCTGTCTGTCTGTG 877
QY 842 TGCTGTCTATCTCAAGCAGCCCTGTGCCAGCGCCGTCCACCTCGAGTGCTCTCTTCAG 901
Db 878 TGCTGTCTATCTCAAGCAGCCCTGTGCCAGCGCCGTCCACCTCGAGTGCTCTCTTCAG 937
QY 902 CCTGTTTCCCAGCCACAGGCTGTCTAGACCCCGCCACGATGTGCCGCAAGCACTGTCTCA 961
Db 938 CCTGTTTCCCAGCCACAGGCTGTCTAGACCCCGCCACGATGTGCCGCAAGCACTGTCTCA 997
QY 962 CCATCCCGACCCACGACAGCCAGCGGTGGAGTCCAGGCCA 1007
Db 998 CCATCCCGACCCACGACAGCCAGCGGTGGAGTCCAGGCCA 1043

RESULT 7
AX014147 AX014147 895 bp DNA linear PAT 07-SEP-2000
LOCUS Sequence 12 from Patent WO9954447.
DEFINITION AX014147
ACCESSION AX014147
VERSION AX014147.1 GI:10040594
KEYWORDS
SOURCE
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 895)
Schmitt,A., Specht,T., Dahl,E., Hinemann,B., Rosenthal,A. and
Pliarsky,C.
Human nucleic acid sequences of bladder tumour tissue
PATENT: WO 9954447-A 12 OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PLIARSKY CHRISTIAN (DE)
FEATURES
location/Qualifiers
1..895
/db_xref="taxon:9606"
BASE COUNT 249 a 205 c 228 g 213 t
ORIGIN

Query Match 26.6%; Score 866.2; DB 6; Length 895;
Best Local Similarity 99.5%; Pred. No. 3.1e-212;
Matches 879; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 2358 AGAACCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2417
Db 14 AGAACCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 73
QY 2418 CCTGACTTACCAAAACAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2477
Db 74 CCTGACTTACCAAAACAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 133
QY 2478 ATAGCTTAGATGGTGAAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2537
Db 134 ATAGCTTAGATGGTGAAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 193
QY 2538 AAAGGAGACAGTCACTCTAGATGAAGCTGTGACTAGGCTGGAGTGTCTCTCTCTGAAGA 2597
Db 194 AAAGGAGACAGTCACTCTAGATGAAGCTGTGACTAGGCTGGAGTGTCTCTCTCTGAAGA 253
```

```
QY 2598 TGGGACTCCTTGGGTATCAAGACCTATGCCATCACACTGGGCTAGGAGTAGGTCA 2657
Db 254 TGGGACTCCTTGGGTATCAAGACCTATGCCATCACACTGGGCTAGGAGTAGGTCA 313
QY 2658 TGCCAGCCCTCAAGTCTGTCTTACCCAGGGACTTTGAGAAGTTATATTGGGCAGTGGCTC 2717
Db 314 TGCCAGCCCTCAAGTCTGTCTTACCCAGGGACTTTGAGAAGTTATATTGGGCAGTGGCTC 373
QY 2718 CAATCTGTGGACCACTATTTAGCTTTCCCTGAAAGATCAGGAGGCTCCATTCATTGTC 2777
Db 374 CAATCTGTGGACCACTATTTAGCTTTCCCTGAAAGATCAGGAGGCTCCATTCATTGTC 433
QY 2778 TTTCTCTCTAGCCCTCAGGAAAGAGGACTATATTGTACTGTACCTAGGGGTTCT 2837
Db 434 TTTCTCTCTAGCCCTCAGGAAAGAGGACTATATTGTACTGTACCTAGGGGTTCT 493
QY 2838 GGAAGGAAACATGGAATCAGGATTTATAGACTGTAGAGCCCTATCCAAAGGGCCAT 2897
Db 494 GGAAGGAAACATGGAATCAGGATTTATAGACTGTAGAGCCCTATCCAAAGGGCCAT 553
QY 2898 GACTGGGAAAGGTATGGAGCAGAGGAGATTTGGGATTTTAGGGTGCAGCTACGCTCA 2957
Db 554 GACTGGGAAAGGTATGGAGCAGAGGAGATTTGGGATTTTAGGGTGCAG-TACGCTCA 612
QY 2958 CCCTAAACTTTTGTGGCTGGGCGATGCTTTGAGGCCAGACTGTTAAGCAGGCTCTGC 3017
Db 613 CCCTAAACTTTTGTGGCTGGGCGATGCTTTGAGGCCAGACTGTTAAGCAGGCTCTGC 672
QY 3018 TGGCCTGTTTACTCTGCTACCACTCTGCACCTGTCTTTGAGACTCCATCCAGCCCCAG 3077
Db 673 TGGCCTGTTTACTCTGCTACCACTCTGCACCTGTCTTTGAGACTCCATCCAGCCCCAG 732
QY 3078 GCAGCCCACTGCTCTCTGAGCCTCCACTATCTCCCTGTGACGGGTGAACCTCGTACTG 3137
Db 733 GCAGCCCACTGCTCTCTGAGCCTCCACTATCTCCCTGTGACGGGTGAACCTCGTACTG 792
QY 3138 TGCTCTGGGTCCATATATGAATTGTGAGCAGGGTTCATCTATTTTAAACACAGATGTTTA 3197
Db 793 TGCTCTGGGTCCATATATGAATTGTGAGCAGGGTTCATCTATTTTAAACACAGATGTTTA 852
QY 3198 CAAATAAAGATTATTTCAAACCCACCAAAAAAAGAAAAA 3240
Db 853 CAAATAAAGATTATTTCAAACCAAAAAAAGAAAAAAGAAAAA 895

RESULT 8
BC027769 3139 bp mRNA linear ROD 07-AUG-2002
LOCUS Mus musculus, clone MGC:38375 IMAGE:5345297, mRNA, complete cds.
DEFINITION BC027769
ACCESSION BC027769
VERSION BC027769.1 GI:20380721
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3139)
Direct Submission
TITLE Strausberg,R.
JOURNAL
National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
```


Db 1625 CGATGGGTGGCAGGAT--TAAGTCACCTCTGTTCTTCTACCTCCCATGCCTC----- 1675
QY 1574 CAATTTCTCTCCACCTCGCTCCCAATTTAAAGATGTGGACCAAGCCCTGTGGTACTCC 1633
Db 1676 ----TCTCTCCTCCTGATTCACAAATTTAAAGATGGGATCGAGTCTGAGGACAC-CA 1730
QY 1634 AGGGCAAGAGAGCCCTCGGGTCACTGACACTGTGAGGCCAACCACTGACCTCCACAAAG 1693
Db 1731 AGGGCAAGGGAGCCCTCGGGTCA--GATACTGTATAGTCAAGTGTGAATCCAGATAG 1788
QY 1694 GGGAGACTTTGAAATGAAGGACTAGCTCTTATGTATCAGTTAAGCAAGGGAGAGCT 1753
Db 1789 AGGCAAGCGGATGCGAAGGGCGGCTCTTGAACACAGGCTGTAGAGCAAGAGCAGAGG 1848
QY 1754 GGCAGGGACAGCAGTTTCACAGCAGAGGGGAATGTAGCAACAGCAGCGGCTCTTAGGC 1813
Db 1849 AGC-----TTCGCAGACAGCAGTCTGTGGGACTACAGGGCATCAGAGT 1896
QY 1814 CCATCTTCCATTTCTTAGGTAAAGAGACATTTCTCAGACTCCAGCGGAGGACTGA 1873
Db 1897 GACAAGTAATC-----AACCACTCCTCTTTCACCTGA 1927
QY 1874 GCTAGCCTTCAGCAACCAAGGTTCTCCTGGGACCCAAAGTTTATGGGAGAGGGCAAG 1933
Db 1928 ACCAGCCTTAGGGAGTTAGGGTTTCCCGAGAGTCTGGAACACAGGCCTTAAGACAAGG 1987
QY 1934 ACTTCATGGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1993
Db 1988 AGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2034
QY 1994 CTTTAAAGACAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2053
Db 2035 CAG-----GACTCCCTCCCTGTTCCCTTATTTCTAAGTCGGCTTTGCCAA 2082
QY 2054 CCACACTGTGGCTACAGGTGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2113
Db 2083 CTGCTGTGGCTACCAAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2142
QY 2114 AAGTTTCTTTAACCCTATATGGCCCAAGAGTAGCTGTAGGAGGAGGAGGAGGAGGAGGAG 2173
Db 2143 AAGTTTCTTTAACC--TGTGGTTCAAGAGTAGCTGTAGGAGGAGGAGGAGGAGGAGGAG 2200
QY 2174 CAAATTAATTA-----CCAGTTCTACTGGGTTCTGCGCCACCGTCCCAAGGTGGCGAG 2228
Db 2201 AAACAACAACAACAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2260
QY 2229 GCTAGAGAGAGGATCTTTAAGCCACACATTAGCTGCTGCTGCTGCTGCTGCTGCTGCT 2288
Db 2261 GCTAGGTAAGAGAGAGTCCCAAGATGCATCTCAAGCCTGGGCTGGAGGCCAAGCAGA 2320
QY 2289 ACAAGAAGCTGGGTGT---TGAGTATTTCATCAACTAAGAAC--AAAATCCAGGCACT 2342
Db 2321 GGAAGAACTAGGCACTGACGAGCACTTCAAGGACCAAGGCTCTTCAATCCAGGCCACA 2380
QY 2343 CATATGTGAAGGATAAGAACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2400
Db 2381 GGTGTGATGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2440
QY 2401 ACCATCAACCTTCTCTCTCTCTTACCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2460
Db 2441 ATCAATGGGAGTTTCTCTCTCTCTTACCAACAGGAGGAGGAGGAGGAGGAGGAGGAGG 2500
QY 2461 GGACTTA 2467
Db 2501 GGATTTA 2507

RESULT 9
LOCUS RATRAB15X 945 bp mRNA linear ROD 27-APR-1993
DEFINITION Sprague-Dawley (clone LRB9) RAB15 mRNA, complete cds.
ACCESSION M83679
VERSION M83679.1 GI:206536

KEYWORDS LMW GTP-binding protein.
SOURCE Rattus norvegicus (strain Sprague-Dawley) (library: LAMBDA ZAPII)
adult brain cDNA to mRNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 945)
AUTHORS Elferink, L.A., Anzai, K. and Schellier, R.H.
TITLE rab15, a novel low molecular weight GTP-binding protein
specifically expressed in rat brain
J. Biol. Chem. 267 (9), 5768-5775 (1992)
MEDLINE 92210533
PUBMED 1313420
FEATURES
Location/Qualifiers
1..945
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
/tissue_lib="LAMBDA ZAPII"
1..945
/gene="RAB15"
220..858
/gene="RAB15"
/codon_start=1
/product="RAB15"
/protein_id="AAA41995.1"
/db_xref="GI:206537"
translation="NAKQYDVLFRLLIGDSGVKGTCLICRFDNEFHSSHLSTIGVD
FMKTIYDGIKVIQIWDTAGERYQITTKOYVRAAGIFLVIYDISERSYOHIMKW
VSDVDYAPEGVKILIGNKADDEKROVREGOOLAKEYGMDFEYFSACNLNIKE
SFRLTELVLQHRKELDLKRLTCAENELALELEDEKTEGPANSSKTCWC"
BASE COUNT 222 a 259 c 292 g 172 t
ORIGIN
Query Match 17.7%; Score 577.6; DB 10; Length 945;
Best Local Similarity 88.7%; Pred. No. 1e-137;
Matches 637; Conservative 0; Mismatches 79; Indels 2; Gaps 1;
QY 21 TTCCGCGCCCGCTGGCCCGAGTCATGCGCAAGCAGTACGATGCTGTTCGCGCTGCTG 80
Db 196 TTCCGCGCCCGCTGGCTCCCTGCTCCATGGCAACAGTACGATGCTGTATTTCCGCTACTG 255
QY 81 CTGATCGGGAGCTCCGGGTGGCAAGACCTGCCTGCTGTCCGCTTCCACGACACGAG 140
Db 256 CTGATCGGGAGCTCCGGGTGGCAAGACCTGCCTGCTATGCGGCTTCCCGCAACGAG 315
QY 141 TTCCACTCTCGCACATCTCCACCATCGGTGTGACTTTAAGATGAAGACCATAGAGGTA 200
Db 316 TTCCACTCTCGCATATCTCCACCATCGGTGTGACTTTAAGATGAAGACCATCGAAGTA 375
QY 201 GACGGCATCAAGTGGGATACAGATCTGGGACACTGCGAGGCGAGGAGATACAGAGCC 260
Db 376 GACGGCATCAAGTGGGATACAGATTTGGGACACAGCAGGCGAGGAGGATACAGAGCT 435
QY 261 ATCACAAGCAGTACTATCGCGGGCCCGAGGATATTTTGGTCTATGACATTACAGC 320
Db 436 ATCACAAGCAGTACTATCGCGGGCCCGAGGAGGATATTTTGGTCTATGACATTACAGC 495
QY 321 GAGCGCTCTTACAGCACATCATGAAGTGGGTGAGTCAAGTGGTCAAGTGGTCAAGCAGCA 380
Db 496 GAGCGCTCTTACAGCACATCATGAAGTGGGTGAGTCAAGTGGTCAAGTGGTCAAGCAGCA 555
QY 381 GCGTCCAGAAGATCCTTATTGGGAATAAGGCTGATGAGGAGCAGAAACCGCAGGTGGGA 440
Db 556 GCGTCCAGAAGATCCTTATCGGGAATAAGGCTGATGAGGAGCAGAAACCGCAGGTGGGG 615
QY 441 AGAGCAGGAGGAGCAGGAGTGGGAGGAGGAGTGGGATGGGATGCTTATGAACAAGTGGCC 500
Db 616 AGAGCAGGAGGAGCAGGAGTGGGAGGAGGAGTGGGATGGGATGCTTATGAACAAGTGGCC 675

33344 33443: gap of unknown length
33444 34975: contig of 1532 bp in length
34975 35075: gap of unknown length
35075 35779: contig of 2504 bp in length
35779 37580: gap of unknown length
37580 39003: contig of 1324 bp in length
39003 39103: gap of unknown length
39103 40564: contig of 1461 bp in length
40564 40665: gap of unknown length
40665 41979: contig of 1315 bp in length
41979 42079: gap of unknown length
42079 43598: contig of 1519 bp in length
43598 45118: contig of 1720 bp in length
45118 47359: contig of 1841 bp in length
47359 47360: gap of unknown length
47360 48918: contig of 1459 bp in length
48918 49018: gap of unknown length
49018 50383: contig of 1365 bp in length
50383 50483: gap of unknown length
50483 52346: contig of 1863 bp in length
52346 52446: gap of unknown length
52446 55257: contig of 2811 bp in length
55257 55357: gap of unknown length
55357 57005: contig of 1648 bp in length
57005 57105: gap of unknown length
57105 59670: contig of 2565 bp in length
59670 59770: gap of unknown length
59770 61952: contig of 2182 bp in length
61952 62052: gap of unknown length
62052 64503: contig of 2451 bp in length
64503 64603: gap of unknown length
64603 65752: contig of 1149 bp in length
65752 65852: gap of unknown length
65852 68166: contig of 2314 bp in length
68166 68266: gap of unknown length
68266 70091: contig of 1825 bp in length
70091 70191: gap of unknown length
70191 72887: contig of 2696 bp in length
72887 74782: contig of 1795 bp in length
74782 74882: gap of unknown length
74882 77248: contig of 2366 bp in length
77248 77348: gap of unknown length
77348 79279: contig of 1931 bp in length
79279 79379: gap of unknown length
79379 83158: contig of 3779 bp in length
83158 83258: gap of unknown length
83258 85363: contig of 2105 bp in length
85363 88527: contig of 3064 bp in length
88527 88628: gap of unknown length
88628 90451: contig of 1824 bp in length
90451 90451: contig of 1824 bp in length

Query Match 13.0%; Score 423.6; DB 2; Length 162495;
Best Local Similarity 61.8%; Pred. No. 1.5e-97;
Matches 1227; Conservative 0; Mismatches 664; Indels 95; Gaps 31;

QY 523 AGTCATTACCGCGCTGACAGACTGGTCTGCAGGCCCATAGAGAGGAGCTGGAAGGCC 582
DB 125590 AGTCCTTCACTCGCTGACCGAGCTGGTGTGCGAGCTCACAGAAAGAGCTGGATGGTC 125531
QY 583 TCCGGATGCGTCAGCAATGAGTTGGCACTGGCAGAGTGGAGAGGAGGGCAAAAC 642
DB 125530 TCCGAACATGTGCCAGCATGAGCTCGACTGGCCGAGCTGGAGGAGAGCAAGGCAAA 125471
QY 643 CCGAGGCCCGACGAACTCTTCGAAACCTGCTGGTGTCTGA--GTCTGTGTGGGGCACC 700
DB 125470 CTGAAGGCCCGACAACTCTTCAAGACCTGCTGGTGTCTGAGGGTCTCTGTGTGAAATCCT 125411
QY 701 CCACAGCACACCCTCTTCCTCAGAGGCCCGCTGGGCGAGACAGGGAGCCGGGGCTTTG 760
DB 1254375 GGTGACAGCAGGGCGGTCCCTGTGAGGACATTTACCCAGACTCCCTGGCAGAGGTACA 124316

DB 125410 CCCTCCCCCAGTCCCTTCCCTCAGGAGGCCCGCTGGAGGCA---GGGAGTCTGGGCTTTG 125354
QY 761 CCCTGCTGTGTCTCTCGTGTGATGACCCCTATTGAGTATCAGTAGCCTACTCTCCCCCT 820
DB 125353 TCCATC-ACTGTCTCTTCAATTGAGATCTCTGTGAAACCACTAGTGTGACT-CCCT 125296
QY 821 GCTGCCCCCTGAGAGCGGCTCTCTGTGTCATCT-CAAGACGCCCTCTGCCCCAGCCGTCC 879
DB 125295 GCCTGGCCCTGTGAGCAGCTTTGCTGTGGTCTCCGAGCAACCTCTGTCCCCAGCCCTGCT 125236
QY 880 ACCCTGGAGTGGTCTTCTTCAGCCTGTTTCCCGAGCCACAGGCGCTGCTAGACCCCCACG 939
DB 125235 ATCCTGGAGTGGTAACCTTTTCCACCTGTTTCCAGCGCGAGGCTGCTGTGAGCCCC---- 125180
QY 940 ATGTGCCGAAGCACTGTCTCACCATCCCGCACCCACAGCAACAGCCAGGCGTGGAGT 999
DB 125179 AGAATGTGCTTCAGCCATCTCACCAACCCAGGCTCTGGGACAAAGGTCAAGGCTGGAGT 125120
QY 1000 CCAGGCCACTTTCAGTGCTCTTCTCCGTGATCGTGTCTCTCTCTCTCTCTCTCTCTCTCT 1059
DB 125119 CAAAGCCACTTAAAGACTCCTTTAGTGCATCTCGTGGGCTCTCTGCTTTTCTCTCTCTCT 125060
QY 1060 CTTCCCCCACTTCTCTTCTCTGACCCCTCCCTCCGGTGGGTTTCTGTATCAAGACTCCT 1119
DB 125059 CTTCTCTTTTCCAGTCCCTTCTTCCATGTATTTTGGCTCCACGTCATCCACGCTCTCT 125000
QY 1120 CAAACCCCGTCCCGTGTGTCTGTGTGAGCTGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCT 1179
DB 124999 C---CTCCCTAGCTGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 124943
QY 1180 TATCCAAAGGGATGGACCCAGGCTGTGGGGAGTTTCCACCTTGGATCCAGGAAGAACC 1239
DB 124942 TGCCCTAAGGAATGGACCCAGGCTC-----TGCCCCCAGGAGGATG 124902
QY 1240 CTCACCCCTGCTGGTGGTGGGCCAAAGGCTACAGGGTCTCTTCTCTCTCTCTCTCTCTCTCT 1299
DB 124901 GCTCACCCCTG-CTCTGGGCTGGGCTGAAGGCTA-AGGGTGTCTTCTCTCTCTCTCTCTCT 124846
QY 1300 CCCACTGTCCCTCATGTGCCATGGGCTGCTGCCAGTGCCTCCCGAGTGCCTGCGAAAGTGGAGCATC 1359
DB 124845 CCTGCTGCCCTTTTGTGCCATGGGCTGCTCCCAATGGCCTG-GGAGGTAGAACAATT 124787
QY 1360 GAGGTAGGAGGAAACAGCAACCCGGGAGTCTCTGAGCCTTGGGGCTGGCCCTACTCTTACC 1419
DB 124786 GAGGTAGGAGGAAACACTATGAGGTG-GTCTCTCAAGCCT-GGACTGGCCTCTCTCTCTCT 124729
QY 1420 CATTCGCCGACAGAGCTTTGCCCTTGTCTGGTCCCGCTGCTCTCTCTCTCTCTCTCTCTCTCT 1479
DB 124728 TTGCCCTTCTCTGAGGTTTGGGCTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 124669
QY 1480 GCTCAGAGCGAGGTGCTTTCAGAGAAGAAACAAAATGAGGGGTGGCAGGATAAAAAGTC 1539
DB 124668 GTCCAGAGCGAGGCACAGCAGAGAGAAGATGACAAGATGGGTGGGGGGGATTAAGTCA- 124609
QY 1540 ACCTCCATTTCTTACCTCCCATCAGCATGAACAATTTCTCTCCACCTGGGCTCCCAAA 1599
DB 124608 --CTCTGTTCTCTTACCTCCCATCAGCGAGTGCACA--GCCCTTTTCAACCGGGTCTCTGG 124553
QY 1600 TTTTAAAGATGTGCACCAAGGCTGTGGTACTCCAGGGGCAAGGAGCCCTGGGGTCAG 1659
DB 124552 ATTTTAAAGATGGGATCAAGGCTTGAAGGCTGAGGAGCGC-CAGAGGGCAAGGGGATCCCCAGGGTCA- 124495
QY 1660 TGACACTGTCCAGCCCAACCATGCATCCCAAAAGGGGAGCATTTTGGAAATGAAGGACTAG 1719
DB 124494 -GACACTGTGTAGTGTGAGCTGTGAGTTAGAGCCACAGCGGATGCGAGAGCTGG 124436
QY 1720 CTCCTATGTATCAGGTTAAGAGCAAGGG-AGAGCTGGGCCAGGAGCAGCTTTTGCACAGC 1778
DB 124435 CTCCTATGTGGCAGGTTAGGACAAGAGTAGAGGGGCTTTCACAGAACAAGTGTGCTGT 124376
QY 1779 AGAGGGGAATGTAGCAACAGCAGGGGCTCTTACGGCCCATCTTCCATTTCTTGTAGGTAAGA 1838
DB 124375 GGTGACAGCAGGGCGGTCCCTGTGAGGACATTTTACCCAGACTCCCTGGCAGAGGTACA 124316

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 15 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1	5033: contig of 5033 bp in length
* 5034	5133: gap of unknown length
* 5134	14755: contig of 9622 bp in length
* 14756	14855: gap of unknown length
* 14955	23282: contig of 8237 bp in length

Db	134265	GC	GGCTCTCTAGAACCAAGG	GTTAGACGAAGAGCAGGGAGC-----TTCCGA	134218
Qy	1775	CAG	CAGAGGGGAAATGTAGCAACAGCAGGGCTCCTAGGCCCATCTTCATCTTCAT	TAGT	1834
Db	134217	GA	CAGCGAGTGTGTGGCGACTACAGGGCATCAGAGGTGCAAGTAATC-----		134168
Qy	1835	AAG	AAGAGCATTTCTCTCAGACTCCCAGGGGAGGACTGAGCCTTAGCTTC	CAGCAACCAAG	1894
Db	134167	-----	-----AACCACCTCTTTCACTGAACCCAGCCTTAGGAGTTAGG		134127
Qy	1895	GT	TCTCTGGGACCCCAAGTTTATGGGAGNAGGGCAAGACACTTCATGGGAAGAGAGAAG		1954
Db	134126	GT	TCCCACAGAGTCTGGCAACAGGGCTTTAAGACAAGGAAGAGAGGGTGGGTA-----		134072
Qy	1955	AAG	CCCTGGGTGAAGACGCTTCGTCTTCCTTTGGCTTTTAAGACAAAGCGCTCAT		2014
Db	134071	-----	-----TGATGGAAATACCTGGTGTCTGGCTTCC-----AGACTCC		134032
Qy	2015	CT	TGCCCTCTACCTCTCTGATAGGCTTTGAGGGTTTGGCAACCACTGTGGCTACAGGTG		2074
Db	134031	CT	CCCTGGTTCCTTATTCTTAAGTCGCTTTTGGCACTGTGTGTGGCTACCAAGG		133972
Qy	2075	AG	GAAGAGAGACTCTCTCTCCAGAGTGCTATGTTCCAGGAAGTTTCTTTAA	CCCATATG	2134
Db	133971	AG	GAAGAGAGACCCCTCTCCAGAGTTCTATCTCAGGAAGTTTCTTTAA	CCCTC	133914
Qy	2135	GCC	CAAGTAGTCTGTAGGAGGCCCTTTAAGACGGACACAGTAATTTA-----C		2185
Db	133913	GT	TCAAGAGTAGCTTGAGGAGGCCCTGAAAACAACAACAACAACAACAACAAGC		133854
Qy	2186	CAG	TTCTACTGGGGTTCTGCCACCGTCCCAGGTGGCGAGGCCCTAGGAAGAGGTCA		2245
Db	133853	CA	CCCTGGTAGGGTTTCATCTTACATCCCAAGTTGGGAGGAGCTCAGTAAGAAGA		133794
Qy	2246	TT	CTTAAGCCACACATATGCTGCACCTGCGTGGCTGCAGCCAAACAAGAACTGGGTGT		2304
Db	133793	GT	CCAAAGATGCATTTCAAGCCTTGGGCTGGAGCCAAAGACAGGAAGAATAGGCAC	TTG	133734
Qy	2305	--	TGAGTATTTCATCACTAAGAACC--AAATCCAGGCCACTCATCTGTAAGGATAG		2359
Db	133733	ACC	GAGCATTCAGAGCAAGGCTTCCTTCAATCCAGGCCAGCAGTGTATGACAGAGAG		133674
Qy	2360	AAC	CTACTTCTTCTTACTCTCTCCAAAAA--GAAGTGGGGAAGAACCATCAAACTTCCT		2417
Db	133673	CCC	ACACTCTTTAGTCTCTTAAGAGGAGATGGGGTGAATCAATGGGATTTCTC		133614
Qy	2418	CCT	GACTTACCAACACAGGAAAAACAGCAGGAGAGGGTGGCTCAGGACTTA		2467
Db	133613	CCT	GACTTAGCAACAAAGGAAGCAGGACGAAGAAGACAGCTCAGGATTTA		133564

[illegible]

ORIGIN

Query Match 9.4%; Score 307; DB 6; Length 313;
 Best Local Similarity 99.0%; Pred. No. 6.1e-68;
 Matches 308; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2824 ACCCTAGGGTTCTGGAGGGAACATGGAAATCAGGATTCATAGACTGATAGGCCCTA 2883
 Db 1 ACCCTAGGGTTCTGGAGGGAACATGGAAATCAGGATTCATAGACTGATAGGCCCTA 60

QY 2884 TCCACAGGGCCATGACTGGGAAAGGTATGGAGCAGAGGAGAAATGGGATTTAGGG 2943
 Db 61 TCCACAGGGCCATGACTGGGAAAGGTATGGAGCAGAGGAGAAATGGGATTTAGGG 120

QY 2944 TGCAGCTACGCTCACCCCTAAACTTTTGGTGGCCCTGGGGCATGTCTTGAGGCCCCAGACTGT 3003
 Db 121 TGCAGCTACGCTCACCCCTAAACTTTTGGTGGCCCTGGGGCATGTCTTGAGGCCCCAGACTGT 180

QY 3004 TAAGCAGGCTCTGCTGGCCCTGTTTACTCGTCACCACTCTGCACCTGCTCTTGAGACT 3063
 Db 181 TAAGCAGGCTCTGCTGGCCCTGTTTACTCGTCACCACTCTGCACCTGCTCTTGAGACT 240

QY 3064 CCATCCAGCCCCAGGCACGCACCTGCTCCTGAGCCCTCCACTATCTCCCTGTGACGGGTG 3123
 Db 241 CCATCCAGCCCCAGGCACGCACCTGCTCCTGAGCCCTCCACTATCTCCCTGTGACGGGTG 300

QY 3124 AACTTCGTGTA 3134
 Db 301 AACTTCGTGTD 311

Search completed: November 17, 2002, 20:03:03
 Job time : 6495 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2002, 20:40:59 ; Search time 1956 Seconds
(without alignments)
1755.339 Million cell updates/sec

Title: US-09-817-198A-2
Perfect score: 1105
Sequence: 1 MAKQYDLFRLLIGDSVG.....LEEEKGPEGPANSSKTCWC 212

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09817198/runat_13112002_135207_7044/app_query.fasta_1.391
-DB=EST -QFMT=fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09817198_ACGN_1_1_899 @runat_13112002_135207_7044 -ICPU=3
-NO_XLPYX -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_Other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	996	90.1	1091	12	BF160330	BF160330 601768601
2	933.5	86.3	932	12	BF535642	BF535642 602054039
3	922	83.4	916	13	BI648588	BI648588 603277781
4	910	82.7	992	12	BF101730	BF101730 601753464
5	910	82.4	902	12	BF178163	BF178163 601809083
6	885	80.1	699	13	BI557933	BI557933 603237549
7	858.5	77.7	796	12	BF966292	BF966292 602286692
8	838	75.8	1100	12	BF181167	BF181167 601805830
9	829	75.0	616	12	BG247902	BG247902 602359819
10	816	73.8	1121	12	BF182001	BF182001 601805936
11	799	72.3	814	12	BG762967	BG762967 602735022
12	792	71.7	660	13	BG974411	BG974411 602844184
13	777	70.3	911	13	BI649317	BI649317 603278051
14	756	68.4	618	12	BF150465	BF150465 uy86b05.y
15	733	66.3	1767	13	BM462147	BM462147 AGENCOURT
16	727	65.8	485	12	BE913096	BE913096 601668580
17	706	63.9	667	13	BI685808	BI685808 603309467
18	704	63.7	843	12	BG865609	BG865609 602783763
19	693.5	62.8	577	12	BG703353	BG703353 602685115
20	686	62.1	465	12	BF565114	BF565114 UI-R-B01-
21	670	60.6	479	12	BF320679	BF320679 uz55d10.y
22	666	60.3	543	12	BG019552	BG019552 daa21e04.y
23	663	60.0	539	12	BF022618	BF022618 uy51f07.y
24	658.5	59.6	693	10	BE287683	BE287683 601093322
25	651	58.9	823	13	BM017178	BM017178 603643780
26	649	58.7	385	14	BM936454	BM936454 UI-W-BH2.
27	589	53.3	956	14	BQ944882	BQ944882 AGENCOURT
28	568.5	51.4	704	13	BJ057692	BJ057692 BJ057692
29	554.5	50.2	714	13	BI259835	BI259835 602971485
30	551.5	49.9	644	14	BQ387194	BQ387194 NISC-mm22
31	547	49.5	611	13	BJ035681	BJ035681 BJ035681
32	547	49.5	678	13	BI395002	BI395002 pppin.pk0
33	547	49.5	788	9	AUI25167	AUI25167 AUI25167
34	545	49.3	613	14	BM791355	BM791355 K-EST0071
35	545	49.3	635	9	AUI34059	AUI34059 AUI34059
36	545	49.3	676	9	AUI33710	AUI33710 AUI33710
37	545	49.3	686	12	BG709122	BG709122 602675387
38	545	49.3	816	9	AUI19700	AUI19700 AUI19700
39	545	49.3	835	12	BG425680	BG425680 602452987
40	545	49.3	860	9	AUI22740	AUI22740 AUI22740
41	545	49.3	877	13	BI763900	BI763900 603049666
42	545	49.3	897	14	BQ671010	BQ671010 AGENCOURT
43	545	49.3	986	13	BM475359	BM475359 AGENCOURT
44	545	49.3	1032	13	BM459427	BM459427 AGENCOURT
45	545	49.3	1059	14	BM923211	BM923211 AGENCOURT

ALIGNMENTS

RESULT 1
BF160330
LOCUS 601768601F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3987745 5',
DEFINITION mRNA sequence.
ACCESSION BF160330
VERSION BF160330.1 GI:11040541
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1091)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: crabbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM9195 row: b column: 02
High quality sequence stop: 654.

	FEATURES	SOURCE
1.	1. The first two columns are labeled "FEATURES" and "SOURCE".	
2.	2. The third column is labeled "FEATURES".	
3.	3. The fourth column is labeled "SOURCE".	
4.	4. The fifth column is labeled "FEATURES".	
5.	5. The sixth column is labeled "SOURCE".	
6.	6. The seventh column is labeled "FEATURES".	
7.	7. The eighth column is labeled "SOURCE".	
8.	8. The ninth column is labeled "FEATURES".	
9.	9. The tenth column is labeled "SOURCE".	
10.	10. The eleventh column is labeled "FEATURES".	
11.	11. The twelfth column is labeled "SOURCE".	
12.	12. The thirteenth column is labeled "FEATURES".	
13.	13. The fourteenth column is labeled "SOURCE".	
14.	14. The fifteenth column is labeled "FEATURES".	
15.	15. The sixteenth column is labeled "SOURCE".	
16.	16. The seventeenth column is labeled "FEATURES".	
17.	17. The eighteenth column is labeled "SOURCE".	
18.	18. The nineteenth column is labeled "FEATURES".	
19.	19. The twentieth column is labeled "SOURCE".	
20.	20. The twenty-first column is labeled "FEATURES".	
21.	21. The twenty-second column is labeled "SOURCE".	
22.	22. The twenty-third column is labeled "FEATURES".	
23.	23. The twenty-fourth column is labeled "SOURCE".	
24.	24. The twenty-fifth column is labeled "FEATURES".	
25.	25. The twenty-sixth column is labeled "SOURCE".	
26.	26. The twenty-seventh column is labeled "FEATURES".	
27.	27. The twenty-eighth column is labeled "SOURCE".	
28.	28. The twenty-ninth column is labeled "FEATURES".	
29.	29. The thirtieth column is labeled "SOURCE".	
30.	30. The thirty-first column is labeled "FEATURES".	
31.	31. The thirty-second column is labeled "SOURCE".	
32.	32. The thirty-third column is labeled "FEATURES".	
33.	33. The thirty-fourth column is labeled "SOURCE".	
34.	34. The thirty-fifth column is labeled "FEATURES".	
35.	35. The thirty-sixth column is labeled "SOURCE".	
36.	36. The thirty-seventh column is labeled "FEATURES".	
37.	37. The thirty-eighth column is labeled "SOURCE".	
38.	38. The thirty-ninth column is labeled "FEATURES".	
39.	39. The fortieth column is labeled "SOURCE".	
40.	40. The forty-first column is labeled "FEATURES".	
41.	41. The forty-second column is labeled "SOURCE".	
42.	42. The forty-third column is labeled "FEATURES".	
43.	43. The forty-fourth column is labeled "SOURCE".	
44.	44. The forty-fifth column is labeled "FEATURES".	
45.	45. The forty-sixth column is labeled "SOURCE".	
46.	46. The forty-seventh column is labeled "FEATURES".	
47.	47. The forty-eighth column is labeled "SOURCE".	
48.	48. The forty-ninth column is labeled "FEATURES".	
49.	49. The fiftieth column is labeled "SOURCE".	
50.	50. The fifty-first column is labeled "FEATURES".	
51.	51. The fifty-second column is labeled "SOURCE".	
52.	52. The fifty-third column is labeled "FEATURES".	
53.	53. The fifty-fourth column is labeled "SOURCE".	
54.	54. The fifty-fifth column is labeled "FEATURES".	
55.	55. The fifty-sixth column is labeled "SOURCE".	
56.	56. The fifty-seventh column is labeled "FEATURES".	
57.	57. The fifty-eighth column is labeled "SOURCE".	
58.	58. The fifty-ninth column is labeled "FEATURES".	
59.	59. The sixtieth column is labeled "SOURCE".	
60.	60. The sixty-first column is labeled "FEATURES".	
61.	61. The sixty-second column is labeled "SOURCE".	
62.	62. The sixty-third column is labeled "FEATURES".	
63.	63. The sixty-fourth column is labeled "SOURCE".	
64.	64. The sixty-fifth column is labeled "FEATURES".	
65.	65. The sixty-sixth column is labeled "SOURCE".	
66.	66. The sixty-seventh column is labeled "FEATURES".	
67.	67. The sixty-eighth column is labeled "SOURCE".	
68.	68. The sixty-ninth column is labeled "FEATURES".	
69.	69. The seventieth column is labeled "SOURCE".	
70.	70. The seventy-first column is labeled "FEATURES".	
71.	71. The seventy-second column is labeled "SOURCE".	
72.	72. The seventy-third column is labeled "FEATURES".	
73.	73. The seventy-fourth column is labeled "SOURCE".	
74.	74. The seventy-fifth column is labeled "FEATURES".	
75.	75. The seventy-sixth column is labeled "SOURCE".	
76.	76. The seventy-seventh column is labeled "FEATURES".	
77.	77. The seventy-eighth column is labeled "SOURCE".	
78.	78. The seventy-ninth column is labeled "FEATURES".	
79.	79. The eightieth column is labeled "SOURCE".	
80.	80. The eighty-first column is labeled "FEATURES".	
81.	81. The eighty-second column is labeled "SOURCE".	
82.	82. The eighty-third column is labeled "FEATURES".	
83.	83. The eighty-fourth column is labeled "SOURCE".	
84.	84. The eighty-fifth column is labeled "FEATURES".	
85.	85. The eighty-sixth column is labeled "SOURCE".	
86.	86. The eighty-seventh column is labeled "FEATURES".	
87.	87. The eighty-eighth column is labeled "SOURCE".	
88.	88. The eighty-ninth column is labeled "FEATURES".	
89.	89. The ninetieth column is labeled "SOURCE".	
90.	90. The ninety-first column is labeled "FEATURES".	
91.	91. The ninety-second column is labeled "SOURCE".	
92.	92. The ninety-third column is labeled "FEATURES".	
93.	93. The ninety-fourth column is labeled "SOURCE".	
94.	94. The ninety-fifth column is labeled "FEATURES".	
95.	95. The ninety-sixth column is labeled "SOURCE".	
96.	96. The ninety-seventh column is labeled "FEATURES".	
97.	97. The ninety-eighth column is labeled "SOURCE".	
98.	98. The ninety-ninth column is labeled "FEATURES".	
99.	99. The hundredth column is labeled "SOURCE".	
100.	100. The hundred and first column is labeled "FEATURES".	
101.	101. The hundred and second column is labeled "SOURCE".	
102.	102. The hundred and third column is labeled "FEATURES".	
103.	103. The hundred and fourth column is labeled "SOURCE".	
104.	104. The hundred and fifth column is labeled "FEATURES".	
105.	105. The hundred and sixth column is labeled "SOURCE".	
106.	106. The hundred and seventh column is labeled "FEATURES".	
107.	107. The hundred and eighth column is labeled "SOURCE".	
108.	108. The hundred and ninth column is labeled "FEATURES".	
109.	109. The hundred and tenth column is labeled "SOURCE".	
110.	110. The hundred and eleventh column is labeled "FEATURES".	
111.	111. The hundred and twelfth column is labeled "SOURCE".	
112.	112. The hundred and thirteenth column is labeled "FEATURES".	
113.	113. The hundred and fourteenth column is labeled "SOURCE".	
114.	114. The hundred and fifteenth column is labeled "FEATURES".	
115.	115. The hundred and sixteenth column is labeled "SOURCE".	

```

FEATURES
Source
1. .1091
Location/Qualifiers
/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:3987745"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
stem cell origin."
/lab_host="DH10B"
Note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT
library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 297 a 320 c 317 g 157 t
ORIGIN
Alignment Scores:
Pred. No.: 6.93e-117 Length: 1091
Score: 996.00 Matches: 198
Percent Similarity: 94.42% Conservative: 5
Best Local Similarity: 92.09% Mismatches: 9
Query Match: 90.14% Indels: 3
DB: 12 Gaps: 0

```

US-09-817-198A-2 (1-212) x BF160330 (1-1091)

Qy	1	MetaLysGlnTyrAspValLeupheArgLeuLeuLeuIleClyAspSerGlyValGly	20
Db	76	ATGCGAAACAGTACGATGCTGTGTTCCGGCTTCTGCTATCGGGACATCCGGGTGGC	135
Qy	21	LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisIleSerThr	40
Db	136	AAGACATGCTGTGTGCCCTTCCACGACAGAGTTCACCTCTCGGATATCTCCAC	195
Qy	41	IleGlyValAspPheLys-MetLysThrIleGluValAspGlyIleLysValArgIleG	60
Db	196	ATCGGTGTTCACITTTACGATGAAGACTATTCGAAGTAGAGCGCATCAAGTCAGAA	255
Qy	60	nIleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTrzArg	80
Db	256	GATTTGGGCACACAGCAGCGGAGAGAGTACCAAGACTATCACAAAGCAGTACTATCGGG	315
Qy	80	gAlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrClnHisIleW	100
Db	316	AGCCACAGGGAATATTTTAGCTTACGACATTAGCAGTGAGCGCTCTATCAGCATATCAT	375
Qy	100	tLys-TrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleG	120
Db	376	GAACGTGGGTTCAGTGAGTACGTACGCTCCAGAGGAGTCCAGAGATCCTTAAC	435
Qy	120	lyAsnLysAlaaspGluGlnLysArgGlnValGlyArgGluGlnGlyGlnLeuA	140
Db	436	GGAAATAGGCTGATGACGACGAGAAACGGCAGGTGGGGAGAGACAGGCGCAGCAGCTTG	495
Qy	140	laLysGluTyrGlyMetAspPheTyrCluThrSerAlaCysThrAsnLeuAsnIleLysC	160
Db	496	CTAAGGAGTAGCGGCATGGACTTCTAGAAACAAGTCGCTGCACCAACCTTAATATTAAAG	555
Qy	160	luSerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGlyL	180

Db	556	AGTCTTCACTCGTCTGACGGAGCTGTGCTGCAGGCCACAGGAAGAGCTGATGGTC	615
Qy	180	euArgMetArgAlaser-AsnGluLeuAlaLeuAlaGluLeuGluGluGluGlylys	199
Db	616	TCGGAACACGTGCGCAGCAACGAGCTCGACTGCGGCGAGGTGCAGGAGGACGAAGCAAA	675
Qy	200	ProGluGlyProAlaAsnSerSerLysThrCysTrpCys	212
Db	676	CCCTGGGGCCCGACGAATCTCAAAGGACTGGCTGGTGC	714
RESULT 2			
LOCUS	BF535642	602054039F1 NCI_CGAP_SG2 Mus musculus cDNA clone	EST 11-DEC-2000
DEFINITION		mRNA sequence.	
ACCESSION	BF535642		
VERSION	BF535642.1	GI:11623010	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 932)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Jeffrey E. Green, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov		
FEATURES	Plate: LHAM9525 row: m column: 09 High quality sequence stop: 685. Location/Qualifiers		
source	1. 932 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:4193456" /clone_lib="NCI_CGAP_SG2" /lab_host="DH10B (T1 phage-resistant)" /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site:1.1001; Site_2: SALL; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."		
BASE COUNT	239 a	250 c	294 g
ORIGIN	149 t		

Alignment Scores:	
Pred. NO.:	1.54e-111
Score:	953.50
Length:	932
Percent Similarity:	Matches: 187
Best Local Similarity:	Conservative: 7
Query Match:	Mismatches: 9
DB:	Indels: 2
	Gaps: 1
	12

US-09-817-198A-2 (1-212) x BF535642 (1-932)

	Qy	1	MetaLaLysGlnTyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerGlyValGly	20
	Dd	145	ATGGCGAAACAGTAGCATGTCTGTTCCGGCTACTGCTGATCGGGGACTCCGGGGTTGGC	204
	Qy	21	LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisIleSerThr	40
	Dd	205	AAGACATCCCTGCTGTGCCGCTTCACCACAAAGAGTTCCTCTCGCATATCTCCACC	264
	Qy	41	IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValargIleGln	60
	Dd	265	ATCGGTGTGACTTTAAAGATAAAGACTATCGATGTAGCGGCATCAAAAGTGAGAATACAG	324

Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."

BASE COUNT 167 a 154 c 179 g 115 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 8.89e-108 Length: 616
Score: 922.00 Matches: 180
Percent Similarity: 98.92% Conservative: 3
Best Local Similarity: 97.30% Mismatches: 2
Query Match: 83.44% Indels: 1
DB: 13 Gaps: 0

US-09-817-198a-2 (1-212) x BI648588 (1-616)

Qy 28 PheThrAspAsnGluPheHisSerSerHisIleSerThrIleGlyValAspPheLysMet 47
|||||
Db 25 TTCACGACAAAGAGTTCACACTCCCTCGCATATCTCCACCATCGTGTGTTAAGATG 84

Qy 48 LysThrIleGluValAspGlyIleLysValArgIleGlnIleTrpAspThrAlaGlyGln 67
|||||
Db 85 AAGACTATCGATGTAGACGGCATCAAGGTGAAATACAGATTTGGGACACACAGGGCAG 144

Qy 68 GluArgTyrGlnThrIleThrLysGlnTyrTyrArgArgAlaGlnGlyIlePheLeuVal 87
|||||
Db 145 GAGAGTACCAGACTATCACAAAGCAGTACTATCGCGAGCCCA-GGAATATTTAGTC 203

Qy 88 TyrAspIleSerSerGluArgSerTyrGlnHisIleMetLysTrpValSerAspValAsp 107
|||||
Db 204 TACGACATTAGCAGTGAGCGCTCCCTATCAGCATATCATGAAGTGGTCACTGACGTGGAT 263

Qy 108 GluTyrAlaProGluGlyValGlnLysIleLeuIleGlyAsnLysAlaAspGluGln 127
|||||
Db 264 GAGTACGCTCCAGAAGGAGTCCAGAAGATCCTTAATTGGGAATAGGCTGATGAAGCAG 323

Qy 128 LysArgGlnValGlyArgGluGlnGlnGlnLeuAlaLysGluTyrGlyMetAspPhe 147
|||||
Db 324 AAACGGCAGGTGGGAGAGAGCAGCGCAGCAGCTGGCTAAGAGTACGGCATGGACTTC 383

Qy 148 TyrGluThrSerAlaCysThrAsnLeuAsnIleLysGluSerPheThrArgLeuThrGlu 167
|||||
Db 384 TACGAAACAAAGTGCCTGCACCACTTAATATAAAGAGTCTTCACTCGTCTGACGGAG 443

Qy 168 LeuValLeuGlnAlaHisArgLysGluLeuGluGlyLeuArgMetArgAlaSerAsnGlu 187
|||||
Db 444 CTGGTGTCTGCAGGCCACAGAAAGAGCTGGATGTCTCCGAAACACGTGCCACACAGAG 503

Qy 188 LeuAlaLeuAlaGluLeuGluGluGluGlyLysProGluGlyProAlaAsnSerSer 207
|||||
Db 504 CTCGCACTGGCGAGCTGGAGGAGGAGCAGCAAAACCTGAGGGCCCAACTCTTCA 563

Qy 208 LysThrCysTrpCys 212
|||||

Db 564 AAGACCTGTGTGTGC 578

RESULT 4
BF101730
LOCUS 601753464F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3981183 5',
DEFINITION mRNA sequence.
ACCESSION BF101730
VERSION BF101730.1 GI:10884256
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 992)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

Qy 61 IleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
|||||
Db 325 ATTTGGACACAGCAGGAGGATACAGACTATCACAAAGCAGTACTATCGCGCA 384

Qy 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
|||||
Db 385 GCCCAGGAATATTTTATGCTACGACATTAGCAGTACGCGCTCCATCAGCATATCATG 444

Qy 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
|||||
Db 445 AAGTGGTACGAGCTGGATGAGTACGCTCCAGAAGGAGTCCAGAAGATCCTAATTGGG 504

Qy 121 AsnLysAlaAspGluGluGlnLysArgGlnValGlyArgGluGlnGlnGlnLeuAla 140
|||||
Db 505 RATAAGGCTGATGAAGACGACAAACGCGAGTGGGGGAGAGCAGCGGCAGCTGGCT 564

Qy 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
|||||
Db 565 AAGGAGTACGGCATGGACTTCTACGAAACAAAGTGCCTCGCAGCAACCTTAATATTCAAGAG 624

Qy 161 SerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGlyLeu 180
|||||
Db 625 TCCTTCACTCGTCTCAGGAGCTGGTGTGACAGGCCACAGAAAGAGCTGGATGGGCTC 684

Qy 181 ArgMetArgAla-SerAsnGluLeuAlaLeuAla---GluLeuGluGluGluGlyLys 199
|||||
Db 685 CGAACACAGTCCCGCAGCAGCTCCACATGGCGGAGAGCTGTGAGAGCGGACGAAAGCGCA 744

Qy 199 sProGluGlyPro 203
|||||

Db 745 ACCTGATGGCCA 757

RESULT 3
BI648588
LOCUS 603277781F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5317917 5',
DEFINITION mRNA sequence.
ACCESSION BI648588
VERSION BI648588.1 GI:15562824
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 616)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11806 row: a column: 22
High quality sequence start: 2
High quality sequence stop: 614.
Location/Qualifiers
1. 616
/organism="Mus musculus"
/strain="129,C57BL/6J,FVB/N"
/db_xref="taxon:10090"
/clone_image="IMAGE:5317917"
/clone_lib="NCI_CGAP Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

FEATURES
source

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM9177 row: p column: 16
High quality sequence stop: 706.

FEATURES

RESULT 5	
BF178163	
LOCUS	902 bp mRNA linear EST 31-OCT-2000
DEFINITION	601809083F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4039701 5',
	mRNA sequence.
ACCSSION	BF178163
VERSION	BF178163.1 GI:11056305
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1. (bases 1 to 902)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

JOURNAL
COMMENT

```

/label=Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
/label=Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
272 c 306 a 163 t
272 c

```

Alignment Scores:		
Pred. No.:	2,01e-106	Length:
Score:	944.00	Matches:
Percent Similarity:	86.85%	Conservative:
Best Local Similarity:	84.98%	Mismatches:
Query Match:	82.71%	Indels:
DB:	12	Gaps:
		1
		992

US-09-817-198A-2 (1-212) x BF101730 (1-992)

QY 1 MetAlaLysGlnTyrAspValLeuPheArgLeuLeuIleGlyAspSerGlyValGly 20
 |||||
 Db 150 ATGCGGAACACAGTACGATGCTGTCCGGCTACTGCTGATCGGGGACATCCGGGTGGC 209

QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisIleSerThr 40
 |||||
 Db 210 AAGACATGGCTGCTGTGCCGGTTCACGCACACAGAGTTCACACTCTCGCATATCTCCACC 269

QY 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyTyrLeLysValArgIleGln 60
 |||||
DB 270 ATCGGTGGTGACTTTAAGATGAAGACATTCGATGTAGACGGCATCAAACTGAGATAAC 329

QY 61 IletrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrIysGlnTyrTyrArgArg 80
 |||||
 Db 330 ATTTTGGGACACAGCAGGGCAGGAGGTACCAAGCTATCACAAGCAGTACTATCGGCGA 389

QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
|||||
Db 390 GCCCAG----- 395

QY 101 LysTrpValSerAspValAspGluTrpAlaProGluGlyValGlnLysIleLeuIleGly 120
pb 395 -----TACGCTCCAGAAGAGCTCCACAGAAGATCCTAATTGGG 431

QY 121 AsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlyGlnLeuAla 140
 |||||
 Db 432 AATRAAGGCTCATGAAGAGCAGAAACGCGAGCTGGGGAGAGAGAGGGGAGCAGCTGGCT 491

QY 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
 492 AAGCAGTACGCATGCGACTTCTACGAACCAAGTCCCTGCACCACTCTAATATTAAAGAG 551
 nb

Qy 161 SerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGlyLeu 180
 |||||
 552 TCCATCAGTCCTGACGGAGTGGCTGCGAGCCCAACAGAAAGAGCTGGATGGCTCTC 611
 |||||

	FEATURES	SOURCE
1	100	100
2	100	100
3	100	100
4	100	100
5	100	100
6	100	100
7	100	100
8	100	100
9	100	100
10	100	100
11	100	100
12	100	100
13	100	100
14	100	100
15	100	100
16	100	100
17	100	100
18	100	100
19	100	100
20	100	100
21	100	100
22	100	100
23	100	100
24	100	100
25	100	100
26	100	100
27	100	100
28	100	100
29	100	100
30	100	100
31	100	100
32	100	100
33	100	100
34	100	100
35	100	100
36	100	100
37	100	100
38	100	100
39	100	100
40	100	100
41	100	100
42	100	100
43	100	100
44	100	100
45	100	100
46	100	100
47	100	100
48	100	100
49	100	100
50	100	100
51	100	100
52	100	100
53	100	100
54	100	100
55	100	100
56	100	100
57	100	100
58	100	100
59	100	100
60	100	100
61	100	100
62	100	100
63	100	100
64	100	100
65	100	100
66	100	100
67	100	100
68	100	100
69	100	100
70	100	100
71	100	100
72	100	100
73	100	100
74	100	100
75	100	100
76	100	100
77	100	100
78	100	100
79	100	100
80	100	100
81	100	100
82	100	100
83	100	100
84	100	100
85	100	100
86	100	100
87	100	100
88	100	100
89	100	100
90	100	100
91	100	100
92	100	100
93	100	100
94	100	100
95	100	100
96	100	100
97	100	100
98	100	100
99	100	100
100	100	100

```

1. .902
Location/Guairalreis
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:4039701"
/clone_lib="NCL CGAP Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="Dhl0B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NYU"

```

	N1H		
BASE COUNT	216 a	247 c	258 g
ORIGIN			
			181 t

Alignment Scores:

Pred. No.:	5.63e-106	Length:	902
Alignment scores:		Matches:	177
Score:	910.00	Conservative:	2
Percent Similarity:	97.28%	Mismatches:	4
Best Local Similarity:	96.20%	Indels:	1
Query Match:	82.35%	Gaps:	0
DR:	12		

US-09-817-198A-2 (1-212) x BF178163 (1-902)

30 AsnAsnGlnpHeHisSerHisTleSerThrTleGlvValAsppheLysMetLysThr 49

14 CACACCCGCTTCCACTCCTCCGCAATATCTCCACCATCGGTGTTGACTTTAAGATGAAGACT 73

50 T l o c l u v a l a s c n c l u r l o r u e v a l a r c l e c l n i l e t r n d s e n t h r a l a c l v g l n c l u a r a 69

74 TCCGACCTCACCCTATCATCTATTTCCTCACCAAGGCAGCG 13
|||||

d6 |||||

Qy 70 TyrGlnThrIleThrLysGlnTyrTyrArgArgAlaGlnGlyIlePheLeuValTyrAsp 89
 Db 134 TACCACTATACAAACAGCTACTATCGCGAGCCAGGGAATATTTTAGTCTACGAC 193

Qy 90 IleSerSerGluArgSerTyrGlnHisIleMetLysTrpValSerAspValAspGluTyr 109
 Db 194 ATTAGCAGTGAAGCGCTCTATCAGCATATCATGAAGTGGTCAAGTGGATGAGTAC 253

Qy 110 AlaProGluGlyValGlnLysIleLeuIleGlyAsnLysAlaAspGluGluGlnLysArg 129
 Db 254 GCTCCAGAGGAGTCCAGAAAGTCTTAATTTGGATAAGGCTGATGAAGACAGAAACGG 313

Qy 130 GlnValGlyArgGluGlnGlnGlnLeuAlalys-GluTyrGlyMetAspPheTyrGln 149
 Db 314 CAGGTGGGAGAGAGAGAGGAGGAGGAGTGGCTAAGCGAGTACGGCATGGACTTCTACGA 373

Qy 149 uThrSerAlaCysThrAsnLeuAsnIleLysGluSerPheThrArgLeuThrGluLeuVa 169
 Db 374 ACAAAGTGCCTGCACCAACCTTAATATTAAAGAGTCTTCACTCGTCTGACGGAGCTGGT 433

Qy 169 IleuGlnAlaHisArgLysGluLeuGluGlyLeuArgMetArgAlaSerAsnGluLeuAl 189
 Db 434 GCTGAGGCCCCACAGGAGAGTGGTCTCGAACAAGTGCACCAAGAGCTCGC 493

Qy 189 aLeuAlaGluLeuGluGluGluGlyLysProGluGlyProAlaAsnSerSerLysTh 209
 Db 494 ACTGGCCGAGCTGGAGGAGGAGGAGGCAAACTTGAGGGCCAGCAAACTCTTCAAAGAC 553

Qy 209 rcYstrpcys 212
 Db 554 CTGCTGGTGC 563

RESULT 6
 BI557933
 LOCUS 603237549f1 NCI_CGAP_Mam3 Mus musculus cdna clone IMAGE:5290284 5',
 DEFINITION mRNA sequence.
 VERSION BI557933
 KEYWORDS EST.
 SOURCE BI557933.1 GI:15445247
 ORGANISM house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapb-r@mail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1A011734 row: b column: 13
 High quality sequence stop: 698.
 Location/Qualifiers
 1. 699
 /organism="Mus musculus"
 /strain="129,C57BL/6J,FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5290284"
 /clone_lib="NCI_CGAP_Mam3"
 /tissue_type="tumor, gross tissue"
 /dev_stage="10 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH

Reference for transgenic model: Xu et al., Nature Genetics
 22, 37-43 (1999)."

BASE COUNT 164 a 188 c 219 g 128 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6.1e-103 Length: 699
 Score: 885.00 Matches: 172
 Percent Similarity: 98.87% Conservative: 3
 Best Local Similarity: 97.18% Mismatches: 1
 Query Match: 80.09% Indels: 1
 DB: 13 Gaps: 0

US-09-817-198A-2 (1-212) x BI557933 (1-699)

Qy 1 MetAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerGlyValGly 20
 Db 156 ATGGCGAAACAGTACGATGTCTGTCCGGCTACTGCTGATCGGGGACTCCGGGTTGGC 215

Qy 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisIleSerThr 40
 Db 216 AAGACATGCTGCTGCTGCCGCTTCACCGACAACGAGTTCCACTCTCGCATATCTCCACC 275

Qy 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGln 60
 Db 276 ATCGGTGTGTGACTTTAAGATGAAGACTATCGATGTAGACGGCATCAAAAGTGAAGAATACAG 335

Qy 61 IleTyrAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
 Db 336 ATTTGGGACACAGCAGGAGGAGAGTACCAGACTATCACAAAGCAGTACTATCGGGGA 395

Qy 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
 Db 396 GCCCAGGGAATATTTTAGTCTACGACATAGCAGTACGCGCTCTATCAGCATATCATG 455

Qy 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
 Db 456 AAGTGGTCACTGAGTGGTGGTACGAGTCCAGAAAGAGTCCAGAAGATCCTAATATGGG 515

Qy 121 AsnLysAlaAspGluGluGlnLysArgGlnVal-clyArgGluGlnGlnGlnLeuAl 140
 Db 516 AATCAGGCTGATGAAGAGCAAAACGGCAGGTGGGGCAGAGAGAGGCGCAGCAGCTGGC 575

Qy 140 aLysGluTyrGlyMetAspPheThrGluThrSerAlaCysThrAsnLeuAsnIleLysGln 160
 Db 576 TAAGAGTACGCGATGGAGTCTTACGAAACAAGTGCCTGCACCAACCTTAATATTAAGA 635

Qy 160 uSerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGlu 176
 Db 636 GTCTTTCACTGCTGTGACGAGAGTGGTGTGTCAGGCGCCACAGACAGACAG 684

RESULT 7
 BF966292
 LOCUS 602286592f1 NIH_MGC_95 Homo sapiens cdna clone IMAGE:4375366 5',
 DEFINITION mRNA sequence.
 VERSION BF966292
 KEYWORDS BF966292.1 GI:12333507
 SOURCE EST.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapb-r@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)


```

QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisIleSerThr 40
Db 220 AAGACATGCTGCTGTGCGCTTCAACGCAACAGAGTCCACTCCTCGCATATCTCCACC 279
QY 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGln 60
Db 280 ATCCGCTGTGACTTTAAGATGAAGACTATCGATAGACGCAATCAAGAGTGAAGATACAG 339
QY 61 IleTrpAspPheAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrArgArg 80
Db 340 ATTGGGACACAGCAGGCGAGGAGGAGTACAGACTATCAAAAGCAGTACTATCGCGGA 399
QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
Db 400 GCCCAGGGAATATTTAGTCTACGACATAGCAGTAGCGCTCCTATCAGCATATCATG 459
QY 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
Db 460 AAGTGGGTGACGTGACGTGATGAGTACGCTCCAGAAGGAGTCCAGAAGATCTTAATGGG 519
QY 121 AsnLysAlaAspGluGlnLysArgGlnVal-GlyArgGluGlnGlnGlnLeuAl 140
Db 520 AATAAGGCTGATGAAGACGCAACGCGAGGTGGGAGCAGAGCGGCGCAGCTGGC 579
QY 140 aLysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsn--IleLys 159
Db 580 TAAGGAGTACGCGATGACTCTACGAACAAAGTGCCTGGAAACCTTAACATTATACAC 639
QY 160 GluSerPheThrArg-LeuThrGluLeuValLeuGlnAlaHisArgLys-GluLeuGlu 179
Db 640 GAGTCTTCACTCTGTGAGGAGCTGTGCTGAGGCCACACAGGAAACGAGCTGGACT 699
QY 179 lyLeuArgMetArgAlaSerAsnGluLeu-----AlaLeuAlaGluL 193
Db 700 GG-----GCTCCGACACAGGTGGCCACAGCAGCAGCGCCGACTGGCGGAAC 747
QY 193 eu-----GluGluGluGluLysProGluGly-----ProAlaAsnSers 207
Db 748 TTGGGAGGACGACGCAAGGAAACCCGAGGGGCGCCGCAACACCTTCCCAAGCA 807
QY 207 er 207
Db 808 CC 809

RESULT 9
BG247902 616 bp mRNA linear EST 13-FEB-2001
LOCUS 602359819f1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4488197 5',
DEFINITION mRNA sequence.
ACCESSION BG247902
VERSION BG247902.1 GI:12757717
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LMAM10334 row: f column: 06
High quality sequence stop: 616.
Location/Qualifiers

FEATURES

```

```

source 1. .616
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4488197"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 165 a 152 c 183 g 116 t
ORIGIN
Alignment Scores: 7.73e-96 Length: 616
Pred. No.: 829.00 Matches: 158
Score: 99.38% Conservative: 3
Percent Similarity: 97.53% Mismatches: 1
Best Local Similarity: 75.02% Indels: 0
Query Match: 12 Gaps: 0
DB:
US-09-817-198A-2 (1-212) x BG247902 (1-616)
QY 51 GluValAspGlyIleLysValArgIleGlnIleTrpAspThrAlaGlyGlnGluArgTyr 70
Db 3 GATGTAGCGGCATCAAAAGTGTAGATACAGATTTGGGACACAGGCGCAGAGGTAC 62
QY 71 GlnThrIleThrLysGlnTyrArgAlaGlnGlyIlePheLeuValTyrAspIle 90
Db 63 CAGACTATCAAAAGCAGTACTATCGCGAGCCAGGGAATATTTAGTCTACGACATT 122
QY 91 SerSerGluArgSerTyrGlnHisIleMetLysTrpValSerAspValAspGluTyrAla 110
Db 123 AGCAGTGGCGCTCTATCAGCATATCATGAAGTGGGTGAGTGGTGGTGGTGGTGGT 182
QY 111 ProGluGlyValGlnLysIleLeuIleGlyAsnLysAlaAspGluGluGlnLysArgGln 130
Db 183 CCAGAGGAGTCCAGAGATCTTAATGGGAATAGGCTGATGAAGACAGCAAGAGCGGAG 242
QY 131 ValGlyArgGluGlnGlnGlnLeuAlaLysGluTyrGlyMetAspPheTyrGluThr 150
Db 243 GTGGGAGAGAGCAGGCGCAGCTGGCTAAGGAGTACGGCATGGCTTCTACGAAACA 302
QY 151 SerAlaCysThrAsnLeuAsnIleLysGluSerPheThrArgLeuThrGluLeuValLeu 170
Db 303 AGTGCCTGCACCAACCTTAATTAAGAGAGTCTTCACTCTGACGAGCTGGTGTCTG 362
QY 171 GlnAlaHisArgLysGluLeuGluGluArgMetArgAlaSerAsnGluLeuAlaLeu 190
Db 363 CAGGCCACAGAAAGAGCTGGATGGTCTCGAACACAGCTGCCAGCAGCAGCTGCACGTG 422
QY 191 AlaGluLeuGluGluGluGluLysProGluGlyProAlaAsnSerSerLysThrCys 210
Db 423 GCCGAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 482
QY 211 TrpCys 212
Db 483 TGGTGC 488
RESULT 10
BF182001 1121 bp mRNA linear EST 31-OCT-2000
LOCUS 601805936f1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4036664 5',
DEFINITION mRNA sequence.
ACCESSION BF182001
VERSION BF182001.1 GI:11060143
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


REFERENCE 1 (bases 1 to 1121)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LAM9312 row: h column: 09
 High quality sequence stop: 712.

FEATURES

source
 1..1121
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:4036664"
 /clone_lib="NCI CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

BASE COUNT 274 a 328 c 340 g 178 t 1 others
 ORIGIN

Alignment Scores:

Pred. No.: 9,29e-94 Length: 1121
 Score: 816.00 Matches: 176
 Percent Similarity: 85.92% Conservativity: 7
 Best Local Similarity: 82.63% Mismatches: 25
 Query Match: 73.85% Indels: 6
 DB: 12 Gaps: 2

US-09-817-198a-2 (1-212) x BF182001 (1-1121)

QY 1 MetAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerGlyValGly 20
 |||||
 Db 196 ATGGGAACACATGACGTGCTGTTCGGCTACTGCTGATCGGGACTCCGGGGTTGC 255
 QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisLeSerThr 40
 |||||
 Db 256 AAGACATGCTGCTGTGCGCTTCACCGACACGAGTTCCACTCTCGCATATCTCCACC 315
 QY 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyLysValArgIleGln 60
 |||||
 Db 316 ATCGGTGTGACTTAAGATGAAGACTATCGAAGTAGACGGCATCAAGATGAGATACAG 375
 QY 61 IleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgAr 80
 |||||
 Db 376 ATTTGGACACAGCAGCGGACGAGCGGTACCGACTATCACAGCAGTACTATCGGGC 435
 QY 80 galaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMe 100
 |||||
 Db 436 AGCCCGGGAATATTTTATGCTAGCATTAGCAGTACGAGCGTCTATCAGCATATCAT 495
 QY 100 tLysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleG 120
 |||||
 Db 496 GAAGTGGGTGACGTGAGTGGATGAGTACGCTCCAGAAGGAGTCCAGAAGATCCTAATNG 555
 QY 120 yAsnLysAlaAspGluGlnGlnLysArgGlnValGlyArgGluGlnGlnGlnLeuAl 140
 |||||
 Db 556 GAATAAGGCTGATGAAGACACAGAACCGGACGAGTGGGGAGAGAGAGCGGCAGAG-CTGGC 614
 QY 140 aLysGluTyrGlyMetAspPheTyrGluThr-SerAlaCysThrAsnLeuAsnIleLysG 160

Db 615 TAAGGAGTACGGCATGGACTTCTACGAAACAGGTCCCTGGAGCAACCTTCATATTCCAC 674
 QY 160 LuSerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArg---LysGluLeuGluG 179
 |||||
 Db 675 GAGTCTTCACTGCTGACGGAGCTGGTCTGCAGGCCCCACAGGCACAGCAGCTGGATG 734
 QY 179 lYleuArgMetArgAlaSerAsnGlu-LeuAlaLeuAlaGluLeuGluGluGly 198
 |||||
 Db 735 GTCCCGGAACACGCGCGCGCAGGAGCTGGCACCCTGGCGCGG---AGCTGGGCGGACGGA 791
 QY 199 LysProGluGluProAlaAsnSerSerLysThr 209
 |||||
 Db 792 CACACGCGCAACCTGTAGGGGCGCGCAGCAAAACC 824
 RESULT 11
 BG762967
 LOCUS 814 bp mRNA linear EST 15-MAY-2001
 DEFINITION 602735022F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4860437 5',
 mRNA sequence.
 ACCESSION BG762967
 VERSION BG762967.1 GI:14073620
 KEYWORDS EST
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 814)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTP
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM1717 row: 1 column: 06
 High quality sequence stop: 629.

FEATURES
 source
 1..814
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4860437"
 /clone_lib="NIH_MGC_49"
 /tissue_type="melanotic melanoma, high MDR (cell line)"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library."

BASE COUNT 223 a 206 c 252 g 133 t
 ORIGIN

Alignment Scores:
 Pred. No.: 8.53e-92 Length: 814
 Score: 799.00 Matches: 161
 Percent Similarity: 96.43% Conservativity: 1
 Best Local Similarity: 95.83% Mismatches: 3
 Query Match: 72.31% Indels: 3
 DB: 12 Gaps: 0

US-09-817-198a-2 (1-212) x BG762967 (1-814)

QY 48 LysThrIleGluValAspGlyIleLysValArgIleGlnIleTrpAspThrAlaGlyGln 67

Db	3	AGATCATAGAGTAGACGGCATCAAGTGGGATACAGATCGGACACTCGAGGCAG	62
Qy	68	GLuAArgTyrGlnThrIleThrLysGlnTyrTyrArgArgAlaGlnGlyIlePheLeuVal	87
Db	63	GAGAGATACCAGACCATCATCAAAAGCAGTACTATCGCGGGCCAGGGGATATTTGGGTC	122
Qy	88	TyrAspIleSerSerGluArgSerTyrGlnHisIleMetLysTrpValSerAspValAsp	107
Db	123	TATGACATTAGCAGCAGGCGCTTTACCAGCACATCATAAGTGGGTCTAGTCAGCTGGAT	182
Qy	108	GLuTyrAlaProGluGlyValGlnLysIleLeuIleGlyAsnLysAlaAspGluGluGln	127
Db	183	GAGTAGCGACCAAGAGCGTCCAGAAGATCCCTTATGGGAAATGAAGCTCATGAGGACGAG	242
Qy	128	LysArgGlnValGlyArgGluGln-GlyGlnGlnLeuAlaLysGluTyrGlyMetAspPh	147
Db	243	AAACGGCAGGTGGGAAGAGACGAAGGGCAGCAGCTGGCGAAGGAGTAGTGCGATCGACTT	302
Qy	147	eTyrGluThrSerAlaCysThrAsnLeuAsnIleLys-GluSerPheThrArgLeuThrG	167
Db	303	CTATGAACAAGTGCCTGCACCAACCTCAACATTAAACAGAGTCATTACGCGCTCTGACAG	362
Qy	167	luLeuValLeuGlnAlaHisArgLysGluLeuGluGlyLeuArgMetArgAlaSer-Asn	186
Db	363	AGCTGTGTCTGCAGGCCCATAGGAAGGAGCTGGGAAGCGCTCCGGATGCGTGCCAGCAAT	422
Qy	187	GluLeuAlaLeuAlaGluLeuGluGluGluGlyLysProGluGlyProAlaAsnSer	206
Db	423	GAGTGGCACTGGCAGAGCTGGAGGAGGAGGAGGGCGCAACCCGAGGGCCCGCAACTCT	482
Qy	207	SerLysThrCysTrpCys	212
Db	483	TCGAAAACCTGCTGGTGC	500

RESULT 12	660 bp	mrna	linear	EST 12-JUN-2001		
LOCUS	602B44184F1	NCI_CGAP_Mam4	Mus musculus	cDNA clone IMAGE:4979781',		
DEFINITION	mRNA sequence.					
ACCESSION	BC974411					
VERSION	BC974411.1	GI:14362048				
KEYWORDS	EST.					
SOURCE	house mouse.					
ORGANISM	Mus musculus					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE	1 (bases 1 to 660)					
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .					
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)					
JOURNAL	Unpublished (1999)					
COMMENT	Contact: Robert Strausberg, Ph.D.					
	Email: cgapbs-r@mail.nih.gov					
	Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth					
	Ph.D.					

FEATURES

```

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo df.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Priscilla Furch,
NIH Reference for transgenic model: Li et al., Cell Growth
and Differentiation 7, 3-11 (1996)."
```

```

BASE COUNT      152 a      186 c      206 g      116 t
ORIGIN
Alignment Scores:
Pred. No.:      4.85e-91      Length:      660
Score:          792.00      Matches:      154
Percent Similarity: 98.11%      Conservative: 2
Best Local Similarity: 96.86%      Mismatches: 2
Query Match:    71.67%      Indels:      1
DB:             13      Gaps:      0
US-09-817-198A-2 (1-212) x BG974411 (1-660)

```

Qy	1	MetalLysGlnTyrrAspValLeupheArgLeuLeuLeuile-GlyAspSerGlyValGl	20
Db	186	ATGGCGAAACAGTACGATGTCTTCGGCCTACTGTGATCTCGGCAGCTCCGGGGTTGG	245
Qy	20	yLysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisIleSerH	40
Db	246	C AAGACATGGCTGTGTGCCGCTTCACCGAACAGAGTTCCTACTCTCGCATATCTCCAC	305
Qy	40	rIleGlyValAspPhelysMetLysThrlieGluValAspGlyIleLysValargIleGl	60
Db	306	CATCGGGTGTTGACTTTACGATACGCACTATCGAAGTAGAGCGCATCAAAGTGAGAATA	365
Qy	60	nIleTrpAspThrAlaGlyGlnGluArgTyrgInThrIleThrLysGlnTyrrYrArgar	80
Db	366	GATTGGGCACACAGCAGGCGAGAGAGTACCAGACTATCACAAAAGCATACTATCGGG	425
Qy	80	qAlaGlnclYlfePheLeuValTyrrAspIleSerSerGluArgSerTyrrGlnHisIleme	100
Db	426	AGCCCAGGGAAATATTTTAGCTACGACATTAGCAGTGAGCGCTCCTATCAGCATATCAT	485
Qy	100	tLysTrpValSerAspValaspGluTyrrAlaProGluGlyValGlnLysIleLeuIleGl	120
Db	486	GAAGTGGGTGACGTACGTGGTACGTACGCTCCAGAAGGAGTCCAGAAGATCCTAAATTGG	545
Qy	120	yAsnLYSAlaspcglucGlnLnylsArgrGlnValVArgGluGlnGlnGlnGlnLeuAl	140
Db	546	GAATAAGCCTGATGAAGACGACGAAAAACGGCAGGTGGGGAGAGACAGGGCGACAGCTGGC	605
Qy	140	aLysGluTyrrcLyMetAspPheTyrrGluThrSerAlaCysThrAsnLeuasnlle	158
Db	606	TCAGAGCTACGGATGACATTCTACGAAACAAGTGCCTGCCCAACTTCATATA	660

RESULT 13				
BI649317				
LOCUS	BI649317	911 bp	mRNA	linear
DEFINITION	603278051F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5318388 5', mRNA sequence.			

BI649317
 BI649317.1 GI:15563553
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 1 (bases 1 to 911)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 Contact: Robert Strausberg, Ph.D.
 COMMENT

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLAMI1807 row: e column: 13
High quality sequence stop: 777.
Location/Qualifiers

FEATURES

1. 911
/organism="Mus musculus"
/strain="129,C57BL/6J,FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5318388"
/clone_lib="NCI_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."

BASE COUNT 217 a 253 c 244 g 195 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 6.8e-89 Length: 911
Score: 777.00 Matches: 157
Percent Similarity: 98.16% Conservativeness: 3
Best Local Similarity: 96.32% Mismatches: 3
Query Match: 70.32% Indels: 2
DB: 13 Gaps: 0

US-09-817-198a-2 (1-212) x BF1649317 (1-911)

QY 50 IleGluValAspGlyIleLysValArgIleGlnIleTyrAspThrAlaGlyGlnGluArg 69
Db 1 ATCGATGTAGAGCGCATCAAGATGAGATACAGATTGGGACACACAGCGAG-GAGAGG 59
QY 70 TyrGlnThrIleThrLysGlnTyrTyrArgArgAlaGlnGlyIlePheLeuValTyrAsp 89
Db 60 TACCAGACTATCAAAAGCAGTACTATCGCGAGGCCA-GGAATATTTTAGTCTACGAC 118
QY 90 IleSerSerGluArgSerTyrGlnHisIleMetLysTyrIleValSerAspValAspGluTyr 109
Db 119 ATTAGCAGTGAGCGCTCCATCAGCATATCATCAAGTGGTCAGTGCATGATGATGATAC 178
QY 110 AlaProGluGlyValGlnLysIleLeuIleGlyAsnLysAlaAspGluGlnLysArg 129
Db 179 GCTCCAGAGAGGAGTCCAGAGATCTTAATGGGAATAGGCTGATGAGAGCAGAAACGG 238
QY 130 GlnValGlyArgGluGlnGlyGlnGlnLeuAlaLysGluTyrGlyMetAspPheTyrGlu 149
Db 239 CAGGTGGGAGAGAGCAGGGGAGCAGCTGGCTAAGGAGTACGGCATGTCACGAA 298
QY 150 ThrSerAlaCysThrAsnLeuAsnIleLysGluSerPheThrArgLeuThrGluLeuVal 169
Db 299 ACAAGTGCCTGCACCAACCTTAATATAAGAGTCTTCACCTGCTGACGAGCTGGTG 358
QY 170 LeuGlnAlaHisArgLysGluLeuGluGlyLeuArgMetArgAlaSerAsnGluLeuAla 189
Db 359 CTGACGGCCCAAGAGAGAGTGGATGTCCTCCGACACAGTCCGACAGCAGCTCGCA 418
QY 190 LeuAlaGluLeuGluGluGluGlyLysProGluGlyProAlaAsnSerSertysThr 209
Db 419 CTGGCGGAGCTGGAGGAGCAAGGCAAACTGAGGGGCCAGCAAACTCTTCAAGAGACC 478
QY 210 CysTrpCys 212
Db 479 TGCTGGTGC 487

RESULT 14
BF150465
LOCUS BF150465 618 bp mRNA linear EST 29-DEC-2000

DEFINITION

uy86b05.y1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3666417 5' similar to SW:RB15_RAT P35289 RAS-RELATED PROTEIN RAB-15. ;, mRNA sequence.

ACCESSION BF150465
VERSION BF150465.1
KEYWORDS GI:11031860
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 618)

REFERENCE

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml

MGI:1427185

Seq primer: -40RP from Gibco
High quality sequence stop: 419.

FEATURES

source

1. 618
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3666417"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"

BASE COUNT 145 a 166 c 194 g 111 t 2 others
ORIGIN

Alignment Scores:

Pred. No.: 1.83e-86 Length: 618
Score: 756.00 Matches: 146
Percent Similarity: 94.81% Conservativeness: 0
Best Local Similarity: 94.81% Mismatches: 8
Query Match: 68.42% Indels: 0
DB: 12 Gaps: 0

US-09-817-198a-2 (1-212) x BF150465 (1-618)

QY 1 MetaLysGlnTyrAspValLeuPheArgLeuLeuIleGlyAspSerGlyValGly 20
Db 157 ATGGCGAAACAGTACGATGCTGTCGCGCTACTCTGATCGGGACTCCGGGGTTGGC 216
QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisLeSerThr 40
Db 217 AAGACATGCTGCTGTCGCGCTTACCAGCAACGAGTCTCCACTCCCGCATATCTCCACC 276
QY 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGln 60
Db 277 ATCGGTGTGACTTAAAGATGAAGACTATCGAAGTAGAGCGCATCAAGTAGAATACAG 336
QY 61 IleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
Db 337 ATTTGGACACAGCAGGGCAGGAGGTACAGACTATCACAAAGCAGTACTATCGCGCA 396
QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100

```
Db 397 GCCCAGGGAATATTNTAGTCTAGCACATTAGCAGTGGCTCTATCAGCATATCATG 456
Qy 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleuIlegly 120
Db 457 AAGTGGGTCACTGACGTGGATGAGTACCTCTCANGAGGAGTCCAGAGATCCTAATTGGG 516
Qy 121 AsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlyGlnLysAla 140
Db 517 AATAATGCTGATGAAGAGCAACCCAGCTGGGGAGAGAGAGGTCACAGCCCTGGC 576
Qy 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThr 154
Db 577 TTAGAGTACGCGATGGACTTCTAGCAACAAAGTGCTGCACC 618

RESULT 15
LOCUS BM462147
DEFINITION AGENCOURT_6424626 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5491242
5', mRNA sequence.
ACCESSION BM462147
VERSION BM462147.1 GI:185111187
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1767)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12111 row: g column: 19
High quality sequence start: 89
High quality sequence stop: 574.
Location/Qualifiers
1..1767
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5491242"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 318 a 704 c 411 g 320 t 14 others
ORIGIN

Alignment Scores:
Pred. No.: 8.62e-83 Length: 1767
Score: 733.00 Matches: 141
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.33% Indels: 0
DB: 13 Gaps: 0

US-09-817-198A-2 (1-212) x BM462147 (1-1767)

Qy 72 ThrIleThrLysGlnTyrTyrArgAlaGlnGlyIlePheLeuValTyrAspIleSer 91
Db 96 ACCATCAAAAGCAGTACTATCGGCGGCCCGGAGATATTTTGGTCTATGACATTAGC 155
Qy 92 SerGluArgSerTyrGlnHisIleMetLysTrpValSerAspValAspGluTyrAlaPro 111
```

```
Db 156 AGCGAGCGCTCTTACCAGCACATCATGAAGTGGGTCACTGACGTGGATGAGTACGCACCA 215
Qy 112 GluGlyValGlnLysIleLeuIleglyAsnLysAlaAspGluGlnLysArgGlnVal 131
Db 216 GAAGCGCTCCAGAGATCCTTATTGGGAATAAGGCTGATGAGGAGCAGAAAACGGCAGGTG 275
Qy 132 GlyArgGluGlnGlyGlnLysGlnLeuAlaLysGluTyrGlyMetAspPheTyrGluThrSer 151
Db 276 GGAAGAGAGCAAGGCGCAGCAGCTGGCGAAGGAGTATGCGCATGGAGCTTCTATGAACAAGT 335
Qy 152 AlaCysThrAsnLeuAsnIleLysGluSerPheThrArgLeuThrGluLeuValLeuGln 171
Db 336 GCCTGCACCAACCTCAACATTAAAGAGTCATTACGCGCTCTGACAGAGCTGGTGTCTGCAG 395
Qy 172 AlaHisArgLysGluLeuGluGlyLeuArgMetArgAlaSerAsnGluLeuAlaLeuAla 191
Db 396 GCCCATAGGAAGGAGCTGGAAGGCTCGCGATGCGCAGCAATGAGTTGGCACTGGCA 455
Qy 192 GluLeuGluGluGluGlyLysProGluGlyProAlaAsnSerSerLysThrCysTrp 211
Db 456 GAGCTGGAGAGGAGGAGGCGCAACCCGAGGCGCGCAACTCTTCGAAAACCTGCTGG 515
Qy 212 Cys 212
Db 516 TGC 518
```

Search completed: November 17, 2002, 22:05:03
Job time : 1962 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model.

Run on: November 17, 2002, 20:03:09 ; Search time 81 Seconds
(without alignments)
539.284 Million cell updates/sec

Title: US-09-817-198A-2
Perfect score: 1105
Sequence: 1 MAKQDVLFRLLIGDSGVG.....LEEBGKPGPANSSKTCWC 212

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rviris.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	798	72.2	168 11 Q91YW0	Q91yw0 mus musculus
2	557	50.4	211 5 Q9RYS2	Q9tys2 caenorhabdi
3	546.5	49.5	204 5 O15971	O15971 drosophila
4	541	49.0	207 11 Q8VCF6	Q8vcf6 mus musculus
5	528.5	47.8	207 5 O18338	O18338 drosophila
6	526.5	47.6	200 4 Q9R0T3	Q9R0t3 homo sapien
7	526	47.6	202 11 Q9DD03	Q9dd03 mus musculus
8	521	47.1	206 3 Q9HET4	Q9het4 aspergillus
9	519	47.0	216 10 Q24466	Q24466 arabidopsis
10	518	46.9	214 10 Q40218	Q40218 lotus japon
11	515.5	46.7	216 10 Q40215	Q40215 lotus japon
12	515	46.6	216 10 Q9FJF1	Q9fjf1 arabidopsis
13	514	46.5	201 5 Q94148	Q94148 caenorhabdi
14	513.5	46.5	215 10 Q40219	Q40219 lotus japon
15	513	46.4	216 10 Q8VWF9	Q8vwf9 nicotiana t
16	512	46.3	216 10 Q40177	Q40177 lycopersico

17	512	46.3	216 10 Q9LZD4	Q9lzd4 arabidopsis
18	511.5	46.3	212 10 Q40217	Q40217 lotus japon
19	510.5	46.2	216 10 Q8W3J4	Q8w3j4 nicotiana t
20	510	46.2	218 10 Q9SF91	Q9sf91 arabidopsis
21	508.5	46.0	215 10 Q41023	Q41023 pisum sativ
22	508	46.0	216 10 Q49844	Q49844 daucus caro
23	508	46.0	216 10 Q8W3J3	Q8w3j3 nicotiana t
24	507.5	45.9	222 10 Q96362	Q96362 brassica ca
25	507	45.9	202 3 Q96VL3	Q96vl3 colletoctic
26	506.5	45.8	216 10 Q9SWV8	Q9swv8 lycopersico
27	506	45.8	216 10 Q41024	Q41024 pisum sativ
28	505	45.7	198 4 Q96G04	Q96gu4 homo sapien
29	505	45.7	216 10 Q8W3J2	Q8w3j2 nicotiana t
30	502.5	45.5	215 10 Q41022	Q41022 pisum sativ
31	495.5	44.8	204 3 Q9C125	Q9clz5 pichia past
32	492.5	44.6	215 10 Q41061	Q41061 pisum sativ
33	484	43.8	205 5 Q9BLF3	Q9blf3 entamoeba h
34	482	43.6	203 10 Q40569	Q40569 nicotiana t
35	482	43.6	205 5 Q26554	Q26554 schistosoma
36	481	43.5	203 10 Q8RU63	Q8ru63 oryza sativ
37	478.5	43.3	201 4 Q9H0U4	Q9hou4 homo sapien
38	478	43.3	202 10 Q9ZRH6	Q9zrh6 petunia nyb
39	477	43.2	202 3 Q9HDT5	Q9ht5 trichoderma
40	477	43.2	202 10 Q39845	Q39845 glycine max
41	477	43.2	202 10 Q08155	Q08155 pisum sativ
42	476	43.1	203 10 Q940Z7	Q940z7 arabidopsis
43	474	42.9	205 10 Q9ZRE2	Q9zre2 arabidopsis
44	473.5	42.9	201 11 Q9D1G1	Q9d1gl mus musculu
45	473.5	42.9	205 5 Q9UAQ6	Q9uaq6 caenorhabdi

ALIGNMENTS

RESULT 1

Q91YW0 PRELIMINARY; PRT; 168 AA.
ID Q91YW0;
AC Q91YW0; (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 19.4 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013790; RAH13790.1;
DR InterPro; IPR001806; Ras_trnsfrmng.
DR InterPro; IPR002078; Sig54_interact.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; Ras; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.
KW GTP-binding; Hypothetical protein.
SQ SEQUENCE 168 AA; 19399 MW; 4E2C2FB1C56BCDF8 CRC64;

Query Match 72.2%; Score 798; DB 11; Length 168;
Best Local Similarity 99.3%; Pred. No. 3.7e-59;
Matches 152; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAKQDVLFRLLIGDSGVGKTCLLCRFTDNEFHSHSHISTIGVDFKMKKTIEVDGIKVRQ 60		
Db	1	MAKQDVLFRLLIGDSGVGKTCLLCRFTDNEFHSHSHISTIGVDFKMKKTIDVDGIKVRQ 60		
QY	61	IWDTAQERYQITTKYYRAQGIFLVYDISSERSQIHKMKVSDVDEYAPGQVKILIG 120		
Db	61	IWDTAQERYQITTKYYRAQGIFLVYDISSERSQIHKMKVSDVDEYAPGQVKILIG 120		
QY	121	NKADEQKQVGREGQQLAKEYGMDFYETSAC 153		

Db	121	NKADEQKRVGREGQQQLAKEYGMDFYETSAC	153
RESULT 2			
Q9TVS2			
ID	Q9TVS2	PRELIMINARY;	PRT; 211 AA.
AC	Q9TVS2		
DT	01-MAY-2000	(TREMBLrel. 13, Created)	
DT	01-JUN-2002	(TREMBLrel. 21, Last sequence update)	
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)	
DE	C. elegans RAB-8 protein (corresponding sequence D1037.4).		
GN	RAB-8.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BRISTOL N2;		
RX	MEDLINE=99069613; PubMed=9851916;		
RA	Waterston R.;		
RT	"Genome sequence of the nematode C. elegans: a platform for		
RT	investigating biology. The C. elegans Sequencing Consortium."		
RL	Science 282:2012-2018(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BRISTOL N2;		
RA	Ledwith J., Biewald T.;		
RT	"The sequence of C. elegans cosmid D1037.4";		
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BRISTOL N2;		
RA	Waterston R.;		
RT	Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.		
RL	EMBL; AF106592; AAK21367.2;		
DR	EMBL; AF106592; AAK21367.2;		
SQ	SEQUENCE 211 AA; 24022 MW; B7609A91B6082DA2 CRC64;		
Query Match 50.4%; Score 557; DB 5; Length 211;			
Best Local Similarity 55.0%; Pred. No. 6.9e-39;			
Matches 104; Conservative 42; Mismatches 39; Indels 4; Gaps 1;			
Qy	1	MAKYDVLFRLLIGDSGVGKTCCLCRFDNEPHSSHISTIGVDFKMTIEVDGKIVRIQ	60
Db	1	MAKYDYLFKLLIGDSGVGKTCVLFREFSDSFNSFISTIGIDFKIRTIELDGKKIKLQ	60
Qy	61	IWDTAGQERYOTIKQYRRAGQFLVYDISSERSYQHIMKWSVDVDEYAPGVQKILIG	120
Db	61	IWDTAGQERFRTITAYYRGAMGILVYDITNERSFENIKWIRNIEHAASDVVERMIIG	120
Qy	121	NKADEQKRVGREGQQQLAKEYGMDFYETSACNLNLIKESFTRLTELVLQAHKKEGL	180
Db	121	NKCDTEERVSRRGEGQLAIEYGTGKFLETSAKANLNIDEAFFTLARDI-----KSKMEQN	176
Qy	181	RMRASNELA	189
Db	177	EMRAGGSVS	185
RESULT 3			
ID	O15971	PRELIMINARY;	PRT; 204 AA.
AC	O15971;		
DT	01-JAN-1998	(TREMBLrel. 05, Created)	
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)	
DE	RAB10 protein (LD39986P).		
GN	RAB10 OR DRAB10 OR CG17060.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		


```
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL: AE003515; AAF49101.1; -.
DR EMBL: D84347; BAA21711.1; -.
DR HSSP: P05713; 3RAB.
DR FlyBase: FBgn0015796; Rab8.
DR InterPro: IPR003579; GTPase_Rab.
DR InterPro: IPR001230; Prenyl_site.
DR InterPro: IPR001806; Ras_trnsmng.
DR InterPro: IPR002078; Sig54_interact.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00071; ras; 1.
DR PRINTS: PR00449; RASTRNSFRMG.
DR SMART: SM00175; RAB; 1.
DR TIGRFAMs: TIGR00231; small_GTP; 1.
DR PROSITE: PS00294; PRENYLATION; UNKNOWN_1.
DR PROSITE: PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.
KW GTP-binding; Lipoprotein.
SQ SEQUENCE 207 AA; 23691 MW; BE9DE812C77DC909 CRC64;

Query Match 47.8%; Score 528.5; DB 5; Length 207;
Best Local Similarity 53.8%; Pred. No. 1.6e-36;
Matches 100; Conservative 41; Mismatches 40; Indels 5; Gaps 1;

QY 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEPHSSHISITIGVDFKMKTEVDGIKVRQ 60
DB 1 MAKTYDYLFRLLIGDSGVGKTCILFRSEDAFNFTTISTIGIDFKIRTELDNKKIKLQ 60

QY 61 IWDTAGQERYQTITKQYRRAOGIFLVYDISSERSYOHIMKWVSDVDEYAPGEGVQKILG 120
DB 61 IWDTAGQERFTITAYRGANGIMLYDITQKSFENIKRWIENENASADVCKMLLG 120

QY 121 NKADEQKRVGRQGOQLAKEYGMDFYETSACTNLNIKESFTRLTVELVQAHKKELEGL 180
DB 121 NKCLTDKRVSKRGQLAEYGIKFMETSAKASINVEEAF-----LTLASDIKAKTEK 175

QY 181 RMASRN 186
DB 176 RMEANN 181

RESULT 6
ID Q9H0T3 PRELIMINARY; PRT; 200 AA.
AC Q9H0T3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 22.5 kDa protein.
GN DAFZP564L1962.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=21154917; PubMed=11230166;
RA Wlemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansoerg W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
RL Genome Res. 11:422-435(2001).
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL: AL136650; CAB66585.1; -.
DR HSSP: P05713; 3RAB.
```

```
DR InterPro: IPR003579; GTPase_Rab.
DR InterPro: IPR001806; Ras_trnsmng.
DR InterPro: IPR002078; Sig54_interact.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00071; ras; 1.
DR PRINTS: PR00449; RASTRNSFRMG.
DR SMART: SM00175; RAB; 1.
DR TIGRFAMs: TIGR00231; small_GTP; 1.
DR PROSITE: PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.
KW GTP-binding; Hypothetical protein; Lipoprotein.
SQ SEQUENCE 200 AA; 22469 MW; 7F01DB88E46EE3EA CRC64;

Query Match 47.6%; Score 526.5; DB 4; Length 200;
Best Local Similarity 55.8%; Pred. No. 2.3e-36;
Matches 96; Conservative 45; Mismatches 30; Indels 1; Gaps 1;

QY 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEPHSSHISITIGVDFKMKTEVDGIKVR 59
DB 1 MAKTYDYLFRLLIGDSGVGKTCVLFRRSDDAFNFTTISTIGIDFKIRTELQGGKIKL 60

QY 60 QIWDTAGQERYQTITKQYRRAOGIFLVYDISSERSYOHIMKWVSDVDEYAPGEGVQKILI 119
DB 61 QIWDTAGQERFTITAYRGANGIMLYDITNGKSFENISKWLRNIDENEDVERMLL 120

QY 120 GNKADEEQKRVGRQGOQLAKEYGMDFYETSACTNLNIKESFTRLTVELVLQ 171
DB 121 GNKCDMDKRVKPGKGQIAREHGIRFFETSAKANINIEKAFLLTAEILR 172

RESULT 7
ID Q9DD03 PRELIMINARY; PRT; 202 AA.
AC Q9DD03;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 0610007N03RIK protein (RIKEN cDNA 0610007N03 gene).
GN 0610007N03RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirni L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nardone P., Ring B., Ringuwald C., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL: AK022303; BAB22000.1; -.
DR EMBL: BC027214; AAB27214.1; -.

```



```
DR HSSP; P05713; 3RAB.
DR MGD; MGI:1915578; 0610007N03RIK.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR001806; Ras_trnsfrmg.
DR InterPro; IPR002078; Sig54_interact.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMS; TIGR00231; small_gtp; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
DR PROSITE; PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.
KW GTP-binding; Lipoprotein.
SQ SEQUENCE 202 AA; 22770 MW; 5DF599432E228AC0 CRC64;

Query Match 47.6%; Score 526; DB 11; Length 202;
Best Local Similarity 54.1%; Pred. No. 2.5e-36;
Matches 92; Conservative 45; Mismatches 33; Indels 0; Gaps 0;

QY 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSHSHISTIGVDFKMTIEVDGIKVRIO 60
DB 1 MAKAYDHLFLLIGDSGVGKTCCLIRFAEDNFNFSTYISTIGIDPKITVDIEGKRILQ 60

QY 61 IWDTAGQERYQTITKQYRRAGIFLVYDIDISSERSYQHIMKWSDVDEYAPEGVQKILIG 120
DB 61 VMDTAGQERFKTITAYYRGAMGILLVYDITDEKSFENIQNMKSIKENASAGVERLLG 120

QY 121 NKADSEKQVREGOQOLAKKEYGMDFYETSACTNLNLIKESFTRLTELVL 170
DB 121 NKCDMEAKQVOREQAELAREHRIFFETSAKSVNVNDEAFSSLDIL 170

RESULT 8
Q9HET4 PRELIMINARY; PRT; 206 AA.
AC Q9HET4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Secretion related GTPase, (SrgA).
GN SRGA.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N402;
RA Punt P.J., Selboth B., Weenink X.O., van Zeijl C.M., Lenders M.,
RA Konetschny C., Ram A.F., Montijn R., Kubicek C.P.,
RA Van Den Hondel C.A.;
RT "Identification and characterisation of a family of secretion related
RT small GTPase encoding genes from the filamentous fungus Aspergillus
RT niger: a putative SEC4 homologue is not essential for growth.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL; AJ278658; CAC17832.1; -.
DR HSSP; P05713; 3RAB.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsfrmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMS; TIGR00231; small_gtp; 1.
KW GTP-binding; Lipoprotein.
SQ SEQUENCE 206 AA; 22823 MW; 501916B795CF8C8C CRC64;

Query Match 47.1%; Score 521; DB 3; Length 206;
Best Local Similarity 60.6%; Pred. No. 6.8e-36;
Matches 97; Conservative 30; Mismatches 33; Indels 0; Gaps 0;
```

```
QY 3 KOYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSHSHISTIGVDFKMTIEVDGIKVRIO 62
DB 5 RNYDFLIKLLIGDSGVGKSCCLLRFSDSFTPTTIGIDFKIRTIELDGKRVKLQIW 64

QY 63 DTAGQERYQTITKQYRRAGIFLVYDIDISSERSYQHIMKWSDVDEYAPEGVQKILIGNK 122
DB 65 DTAGQERFTITAYYRGAMGILLVYDIDERSFNIRTFNSVDEQHASGVHKILIGNK 124

QY 123 ADEQKQVREGOQOLAKKEYGMDFYETSACTNLNLIKESF 162
DB 125 CDWEKRAVSTEQGOQLANELGIPPEVSAKNNINIEKAF 164

RESULT 9
O24466 PRELIMINARY; PRT; 216 AA.
AC O24466;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ATRAB8 (GTPase ATRAB8).
GN RAB8 OR F4P12.310.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Bischoff F., Palme K.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Bloeker H., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X., Quetier F.,
RA Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL; U82434; AAB65088.1; -.
DR HSSP; P05713; 3RAB.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsfrmg.
DR InterPro; IPR002078; Sig54_interact.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMS; TIGR00231; small_gtp; 1.
DR PROSITE; PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.
KW GTP-binding; Lipoprotein.
SQ SEQUENCE 216 AA; 23939 MW; 1648E45B29D4EEB6 CRC64;

Query Match 47.0%; Score 519; DB 10; Length 216;
Best Local Similarity 51.9%; Pred. No. 1.1e-35;
Matches 98; Conservative 42; Mismatches 43; Indels 6; Gaps 2;

QY 5 YDVLFRLLIGDSGVGKTCCLCRFTDNEFHSHSHISTIGVDFKMTIEVDGIKVRIO 64
DB 12 YDLIRLLIGDSGVGKSCCLLRFSDGFTTIGIDFKIRTIELDGKRIQIWD 71

QY 65 AGQERYQTITKQYRRAGIFLVYDIDISSERSYQHIMKWSDVDEYAPEGVQKILIGNK 124
DB 72 AGQERFTITAYYRGAMGILLVYDIDESSFNIRNIRNIEQHASDVNKLGVNKAD 131

QY 125 -EQKQVREGOQOLAKKEYGMDFYETSACTNLNLIKESFTRLTELVLQA----HRKELE 178
DB 132 MDESKRAVPSKSGALADEYGMKFFETSAKTNLNVEEFFSTAKDIKORLADTDARAEPQ 191
```


CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR EMBL; AB015475; BAB08351.1; -.
DR HSSP; P05713; 3RAB.

DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsfrmng.
DR InterPro; IPR002078; Sig54_interact.
DR InterPro; IPR005225; Small_GTP.

DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.

DR SMART; SM00175; RAB; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.

DR PROSITE; PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.

KW GTP-binding; Lipoprotein.

SQ SEQUENCE 216 AA; 23834 MW; E1370672F6EF3364 CRC64;

Query Match 46.6%; Score 515; DB 10; Length 216;

Best Local Similarity 47.6%; Pred. No. 2.3e-35;

Matches 101; Conservative 43; Mismatches 52; Indels 16; Gaps 3;

QY 5 YDVLFRLLIGDSGVGKTCLLCRFTDNEFHSHSHSTIGVDFKMKTIYDGIKVRQIWDI 64

Db ||| : ||| ||| ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| ||| ||| ||| |||

12 YDLIKLLIGDSGVGKSCLLRFSDSGSTTSFTTIGIDFKIRITELDGKRIKLIQIWDI 71

QY 65 AQERYQITKQYRRAGQIFLYVDISSERSYQHIMKWSVDVEYAPEGVQKILGNKAD 124

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

72 AGQERFTITAYYRGAMGILLVYDVTDESSFNIRNIRNEQHASDNVKNILVGNKAD 131

QY 125 -EQKRVGREGOGOLAKEYGMDFYETSACTNLNLIKESFTRLTELVLQA-----HRKELE 178

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

132 MDESRAVPKSGQALADEYGIKFETSAKTNLNVVEFFSIADIKORLADTDSRAEPA 191

QY 179 GLMRASNELALAELEEGKPEGPANSSKTC 210

Db ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

192 TIKISQTDQAGA-----GQATQKSAC 213

RESULT 13

Q94148

ID Q94148 PRELIMINARY; PRT; 201 AA.

AC Q94148;

DT 01-FEB-1997 (Tremblrel. 02, Created)

DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE RAB10-like.

GN T23H2.5.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Wansley P., Bradshaw H.;

RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 1-124 FROM N.A.

RA Nonet M.L., Staunton J.E., Kilgard M.P., Fergestad T., Jorgensen E.,

RA Hartweig E., Horvitz H.R., Meyer B.J.;

RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.

DR EMBL; U80033; AAC48200.1; -.

DR HSSP; U68257; AAB16972.1; -.

DR SMART; SM00175; RAB; 1.

DR TIGRFAMs; TIGR00231; small_GTP; 1.

DR PROSITE; PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.

KW GTP-binding; Lipoprotein.

SQ SEQUENCE 201 AA; 22712 MW; 2D205ABF751EBF1A CRC64;

Query Match 46.5%; Score 514; DB 5; Length 201;

Best Local Similarity 46.9%; Pred. No. 2.5e-35;

Matches 100; Conservative 48; Mismatches 51; Indels 14; Gaps 3;

QY 1 MAKQ-YDVLFRLLIGDSGVGKTCLLCRFTDNEFHSHSHSTIGVDFKMKTIYDGIKVRQI 59

Db ||| : ||| ||| ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| ||| ||| ||| |||

1 MARRPYDMLFKLLIGDSGVGKTCILYRFSDDAENTFTSTIGIDFKIKTIELKGKKIKL 60

QY 60 QIWDTAGQERYQITKQYRRAGQIFLYVDISSERSYQHIMKWSVDVEYAPEGVQKILI 119

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

61 QIWDTAGQERFHTITTSYRGAMGILLVYDITNAKSFNDIAKLNIDSHASEDVYVKMIL 120

QY 120 GNKADEQKRVGREGOGOLAKEYGMDFYETSACTNLNLIKESFTRLTELVLQAHRKELEG 179

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

121 GNKCDMSDRVSVRERGEKIAQDHGISPHETSAKNLNVHVDYFADLAELAIL----- 171

QY 180 LMRASNELALAELEEGKPEGPANSSKTCWC 212

Db ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

172 AKMPDSTD---EQSRDVTNPVQPQOSSSGC 200

RESULT 14

Q40219

ID Q40219 PRELIMINARY; PRT; 215 AA.

AC Q40219;

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-JUN-1996 (Tremblrel. 01, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE RAB8E.

GN RAB8E.

OS Lotus japonicus.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus.

OX NCBI_TaxID=34305;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-ROOT NODULES;

RA MEDLINE-97231679; PubMed-9076991;

RA Borg S., Brandstrup B., Jensen T.J., Poulsen C.;

RT "Identification of new protein species among 33 different small GTP-

binding proteins encoded by cDNAs from Lotus japonicus, and expression

of corresponding mRNAs in developing root nodules.;"

RL Plant J. 11:237-250(1997).

CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.

DR EMBL; Z73948; CAA98176.1; -.

DR HSSP; P05713; 3RAB.

DR InterPro; IPR003579; GTPase_Rab.

DR InterPro; IPR001806; Ras_trnsfrmng.

DR InterPro; IPR002078; Sig54_interact.

DR InterPro; IPR005225; Small_GTP.

DR Pfam; PF00071; ras; 1.

DR SMART; PR00449; RASTRNSFRMNG.

DR TIGRFAMs; TIGR00231; small_GTP; 1.

DR PROSITE; PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.

KW GTP-binding; Lipoprotein.

SQ SEQUENCE 215 AA; 23715 MW; E7912846F919D608 CRC64;

Query Match

Best Local Similarity 46.5%; Score 513.5; DB 10; Length 215;

Matches 100; Conservative 43; Mismatches 53; Indels 17; Gaps 3;

QY 5 YDVLFRLLIGDSGVGKTCLLCRFTDNEFHSHSHSTIGVDFKMKTIYDGIKVRQIWDI 64

Db ||| : ||| ||| ||| ||| ||| : ||| : ||| ||| ||| ||| ||| : ||| ||| ||| ||| |||

12 YDLIKLLIGDSGVGKSCLLRFSDSGSTTSFTTIGIDFKIRITELDGKRIKLIQIWDI 71

QY 65 AQERYQITKQYRRAGQIFLYVDISSERSYQHIMKWSVDVEYAPEGVQKILGNKAD 124

Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

72 AGQERFTITAYYRGAMGILLVYDVTDESSFNIRNIRNEQHASDNVKNILVGNKAD 131

QY 125 -EEQKRVGREGQQQLAKYGMDFYETSACTNLNLIKESFTRLTELVLQ-----AHRKEL 177
Db 132 MDESKRAVPTSKQALADEYGIKFFETSAKTNLNVVEVFFSIARDIKQRLADTDSKTEP 191
QY 178 EGLRMRASNELALAELEEEGKPEGPANSSKTC 210
Db 192 TGIKIN-----PDKGSAGEAAQKSACC 214

RESULT 15

OBVWF9 PRELIMINARY; PRT; 216 AA.
AC Q8VWF9;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Ras-related protein RAB8-4 (Ras-related protein RAB8-2).
GN RAB8-4 OR RAB8-2.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRIGHT YELLOW 2;
RA Torimoto N., Shimada K., Ito K., Yamamoto K.;
RT "Characterization of Rab8 from tobacco BY-2 cell."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB079023; BAB84325.1; -.
DR EMBL; AB079021; BAB84323.1; -.
DR InterPro: IPR003579; GTPase_Rab.
DR InterPro: IPR001806; Ras_trnsmng.
DR InterPro: IPR002078; Sig54_interact.
DR Pfam: PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR PROSITE; PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.
SQ SEQUENCE 216 AA; 23971 MW; D5E87FABD0D49E13 CRC64;

Query Match 46.4%; Score 513; DB 10; Length 216;
Best Local Similarity 46.7%; Pred. No. 3.4e-35;
Matches 99; Conservative 44; Mismatches 53; Indels 16; Gaps 3;
QY 5 YDVLFRLLIGDSGVGKTCLLCRFTDNEFHSHSHISTIGVDFKMKTIKVIQIWD 64
Db 12 YDVLKILLIGDSGVGKSCLLRFSDGFTTFTTIGIDFKRTIETLDGKRIKLIWD 71
QY 65 AGQERTQTTKQYRRAQGFILYDISSERSYOHIMKWSDVDVEYAPGVQKILIGNKAD 124
Db 72 AGQERFTTITAYIRGAMGILLYDVTDESSFNIRNIRNIEQHASDNVKNKILVGNKAD 131
QY 125 -EEQKRVGREGQQQLAKYGMDFYETSACTNLNLIKESFTRLTELVLQ-----HRKELE 178
Db 132 MDESKRAVPTSKQALADEYGIKFFETSAKTNLNVVEVFFSIARDIKQRLSEDSKTEP 191
QY 179 GLRMRASNELALAELEEEGKPEGPANSSKTC 210
Db 192 AIRNQSDQAGTS-----GQAAQKSSC 213

Search completed: November 17, 2002, 20:40:55
Job time : 83 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2002, 18:15:09 ; Search time 25 Seconds

(without alignments)
351.719 Million cell updates/sec

Title: US-09-817-198a-2

Perfect score: 1105

Sequence: 1 MAKQYDLVRLLLIGDSGVG.....LEEKGKPEGPANSSKTCWC 212

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	1077	97.5	212	1	RB15_RAT	P35289 rattus norv
2	345	49.3	207	1	RAB8_HUMAN	P24407 homo sapien
3	539.5	48.8	210	1	RAB8_DISOM	P22128 discopoyge o
4	532	48.1	203	1	RB13_HUMAN	P51153 homo sapien
5	531.5	48.1	200	1	RAO1_DISOM	P22127 discopoyge o
6	530.5	48.0	200	1	RB10_CANFA	P24409 canis famli
7	530.5	48.0	200	1	RB10_HUMAN	O88386 homo sapien
8	528.5	47.8	207	1	R88B_HUMAN	Q92930 homo sapien
9	528.5	47.8	207	1	R88B_RAT	P70550 rattus norv
10	526	47.6	206	1	RAB8_MOUSE	P55258 mus musculu
11	510	46.2	200	1	VPT2_SCHPO	P17609 schizosach
12	509	46.1	215	1	RAB1_BETVU	Q39433 beta vulgar
13	508	46.0	216	1	ARA3_ARATH	P28186 arabidopsis
14	505.5	45.7	217	1	VPT2_VOLCA	P36861 volvox cart
15	502.5	45.5	200	1	RB10_RAT	P35281 rattus norv
16	502	45.4	203	1	RYL1_YARLI	P41924 yarrowia li
17	483.5	43.8	208	1	SAS1_DICDI	P20790 dictyosteli
18	483	43.7	208	1	YPT1_MAIZE	P16976 zea mays (m
19	480	43.4	202	1	RIC1_ORISA	P40392 oryza sativ
20	479	43.3	203	1	SAS2_DICDI	P20791 dictyosteli
21	477.5	43.2	201	1	RB35_HUMAN	Q15286 homo sapien
22	477.5	43.2	203	1	YPT1_CHLEP	Q39571 chlamydomon
23	477	43.2	201	1	YPT1_PHYIN	Q01890 phytophor
24	476	43.1	258	1	ARA5_ARATH	P28188 arabidopsis
25	475.5	43.0	203	1	YPT1_VOLCA	P31584 volvox cart
26	475	43.0	203	1	YPT1_NEUCR	P33723 neurospora
27	474.5	42.8	210	1	SEC4_CANAL	O14462 candida alb
28	470.5	42.6	203	1	YPT2_MAIZE	Q05737 zea mays (m
29	470.5	42.6	215	1	SEC4_YEAST	P07560 saccharomyc
30	469	42.4	201	1	RB1B_RAT	P10536 rattus norv
31	469	42.4	205	1	RAB1_LYMST	Q05974 lymanaea sta
32	468	42.4	205	1	RB1A_RAT	P05711 rattus norv
33	467	42.3	203	1	YPT1_SCHPO	P11620 schizosach

ALIGNMENTS

RESULT 1

ID	RB15_RAT	STANDARD;	PRT;	212 AA.
AC	P35289;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DE	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Ras-related protein Rab-15.			
GN	RAB15.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Sprague-Dawley; TISSUE=Brain;			
RX	MEDLINE=92210533; PubMed=1313420;			
RA	Elferink L.A., Anzai K., Scheller R.H.;			
RT	"Rab15, a novel low molecular weight GTP-binding protein specifically expressed in rat brain."			
RL	J. Biol. Chem. 267:5768-5775(1992).			
RL	[2]			
RP	ERRATUM.			
RX	MEDLINE=93054572; PubMed=1429617;			
RA	Elferink L.A., Anzai K., Scheller R.H.;			
RL	J. Biol. Chem. 267:22693-22693(1992).			
CC	-!- FUNCTION: MAY ACT IN CONCERT WITH RAB3A IN REGULATING ASPECTS OF SYNAPTIC VESICLE MEMBRANE FLOW WITHIN THE NERVE TERMINAL.			
CC	-!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN NEURAL TISSUES.			
CC	-!- TISSUE SPECIFICITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M83679; AAA41995.1; -			
DR	PIR; F42148; F42148.			
DR	HSSP; P05713; 3RAB.			
DR	InterPro; IPR003579; GTPase_Rab.			
DR	InterPro; IPR001806; Ras_trnsfmg.			
DR	InterPro; IPR005225; Small_GTP.			
DR	Pfam; PF00071; ras; 1			
DR	PRINTS; PR00449; RASTRNSFRMG.			
DR	SMART; SM00175; RAB; 1.			
DR	TIGRFAMs; TIGR00231; small_GTP; 1.			
KW	GTP-binding; Lipoprotein; Prenylation; Protein transport.			
FT	NP_BIND 15 22			
FT	NP_BIND 63 67			
FT	NP_BIND 121 124			
FT	LIPID 210 210			
FT	LIPID 212 212			
FT	SEQUENCE 212 AA; 24283 MW; 04817DDA66CADE12 CRC64;			

P11476 homo sapien
P22125 discopoyge o
P01123 saccharomyc
Q63482 rattus norv
Q9ulw5 homo sapien
P51156 rattus norv
P25228 drosophila
O95716 homo sapien
P11023 bos taurus
Q63942 rattus norv
P03713 mus musculu
P20336 homo sapien

```
Query Match          97.5%; Score 1077; DB 1; Length 212;
Best Local Similarity 97.6%; Pred. No. 6.9e-80;
Matches 207; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAKQYDVLFRLLIGDSGVGKTCLLCRFTDNEPHSHSHISTIGVDFKMKTIYVDGKIKVRIQ 60
Db 1 MAKQYDVLFRLLIGDSGVGKTCLLCRFTDNEPHSHSHISTIGVDFKMKTIYVDGKIKVRIQ 60

QY 61 IWDTAGQERYOTITKQYRRAGQIFLYVDISSERTSYQHIMKMWSDVDYVAPEGVQKILIG 120
Db 61 IWDTAGQERYOTITKQYRRAGQIFLYVDISSERTSYQHIMKMWSDVDYVAPEGVQKILIG 120

QY 121 NKADEQKRVQREGQOQLAKKEYGMDFYETSACTNINIKESFTRRLTELVLQAHKKEGLE 180
Db 121 NKADEQKRVQREGQOQLAKKEYGMDFYETSACTNINIKESFTRRLTELVLQAHKKEGLE 180

QY 181 RMRASNE 187
Db 181 SPOGSNQ 187

RESULT 3
RAB8_DISOM
ID RAB8_DISOM STANDARD; PRT; 210 AA.
AC P22128;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Ras-related protein Rab-8 (ORA2).
OS Discopryoma ommata (Electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squalea; Hypnosqualea; Pristiorajae; Batoidea;
OC Torpediniformes; Narcinoidei; Narcinidae; Discopyge.
OX NCBI_Taxid=7785;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Electric lobe;
RX MEDLINE=91115900; PubMed=1899244;
RA Ngsee J.K., Elferink L.A., Scheller R.H.;

between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; X56741; CAA40065.1; -.
DR EMBL; S53268; AAB19681.1; -.
DR EMBL; BC002977; AAH02977.1; -.
DR EMBL; X56385; CAB56776.1; -.
DR PIR; B36364; B36364.
DR PIR; B49647; B49647.
DR HSSP; P05713; 3RAB.
DR Genew; HGNC:7007; MEL.
DR MIM; 165040; -.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001230; Prenyl_Site.
DR InterPro; IPR001806; Ras_Tnsfrmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRFS; TIGR00231; small_GTP; 1.
KW GTP-binding; Prenylation; Lipoprotein; Protein transport;
KW Proto-oncogene.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 121 124 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 204 204 GERANYL-GERANYL (BY SIMILARITY).
FT CONFLICT 177 183 LEGNSPQ -> WKATAP (IN REF. 2).
SQ SEQUENCE 207 AA; 23668 MW; AA52DBF54A2CD056 CRC64;

Query Match          49.3%; Score 545; DB 1; Length 207;
Best Local Similarity 52.9%; Pred. No. 4.2e-37;
Matches 99; Conservative 45; Mismatches 43; Indels 0; Gaps 0;

QY 1 MAKQYDVLFRLLIGDSGVGKTCLLCRFTDNEPHSHSHISTIGVDFKMKTIYVDGKIKVRIQ 60
Db 1 MAKQYDVLFRLLIGDSGVGKTCVLFSEDAENSTFSTIGIDFKIRITELDGKRIKIQ 60

QY 61 IWDTAGQERYOTITKQYRRAGQIFLYVDISSERTSYQHIMKMWSDVDYVAPEGVQKILIG 120
Db 61 IWDTAGQERYOTITKQYRRAGQIFLYVDISSERTSYQHIMKMWSDVDYVAPEGVQKILIG 120

QY 121 NKADEQKRVQREGQOQLAKKEYGMDFYETSACTNINIKESFTRRLTELVLQAHKKEGLE 180
Db 121 NKCDVNDKRVSKERGEKALDYGKFMETSAKANINVENAFPTLARDIKAKMDKKLEGN 180

QY 181 RMRASNE 187
Db 181 SPOGSNQ 187

RESULT 3
RAB8_DISOM
ID RAB8_DISOM STANDARD; PRT; 210 AA.
AC P22128;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Ras-related protein Rab-8 (ORA2).
OS Discopryoma ommata (Electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squalea; Hypnosqualea; Pristiorajae; Batoidea;
OC Torpediniformes; Narcinoidei; Narcinidae; Discopyge.
OX NCBI_Taxid=7785;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Electric lobe;
RX MEDLINE=91115900; PubMed=1899244;
RA Ngsee J.K., Elferink L.A., Scheller R.H.;
```

RT "A family of ras-like GTP-binding proteins expressed in electromotor
RL J. Biol. Chem. 266:2675-2680(1991).
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M38391; AAA49232.1; -
DR PIR; B38625; B38625.
DR HSP; P05713; 3RAB.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR001806; Ras_trnsfrmng.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRfams; TIGR00231; small_GTP; 1.
KW GTP-binding; Prenylation; Lipoprotein.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 121 124 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 207 207 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 210 AA; 24164 MW; 0F73EDB0DB8B9EEA CRC64;

Query Match 48.8%; Score 539.5; DB 1; Length 210;
Best Local Similarity 47.4%; Pred. No. 1.2e-36;
Matches 99; Conservative 56; Mismatches 45; Indels 9; Gaps 2;

QY 1 MAKQYDVLFRLLLLIGDSGVGKTCCLCRFTDNEFHSHSTIGVDFKMTIEVDGKVRQ 60
Db 1 MAKTYDYLKLLLLIGDSGVGKTCCLFRSEDAFTTFTIGIDFKIRTVLDGKKIKLQ 60

QY 61 IWDTAGQERYQTITKYRRAGQIFLVYDISERSYQHIMKWSDVDEYAPGVOKILIG 120
Db 61 IWDTAGQERFRTITTYRRGAMGIMKYVDITNEKSFNKNIRNIEEHASDDVERMILG 120

QY 121 NKADEQKRVGREGQOOLAKYGMDFVETSACTNLNINKESFTRTELVLQAHRELEGL 180
Db 121 NKCDMEARKVKQEQADKLAREHGIRFFETSASKSMNVNDEAFSSLDIL-----LKSG 174

QY 181 RMRASNELALAELEEGKPEGPANSSK 208
Db 178 -----NENSLQEAVDKLSPPKKPSQKK 201

RESULT 4
ID RB13_HUMAN STANDARD; PRT; 203 AA.
AC P51153;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ras-related protein Rab-13.
GN RB13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94124602; PubMed=8294494;
RA Zahraoui A., Joberty G., Arpin M., Fontaine J.J., Hellio R.,
RA Tavitian A., Loubard D.;
RT "A small rab GTPase is distributed in cytoplasmic vesicles in non
RT polarized cells but colocalizes with the tight junction marker ZO-1

RT in polarized epithelial cells.";
RL J. Cell Biol. 124:101-115(1994).
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COULD PARTICIPATE IN POLARIZED TRANSPORT, IN THE
CC ASSEMBLY AND/OR THE ACTIVITY OF TIGHT JUNCTIONS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC TIGHT JUNCTIONS OR ASSOCIATED
CC WITH VESICLES SCATTERED THROUGHOUT THE CYTOPLASM IN CELLS LACKING
CC TIGHT JUNCTIONS.
CC -1- TISSUE SPECIFICITY: VARIETY OF EPITHELIA, INCLUDING INTESTINE,
CC KIDNEY, LIVER, AND OF ENDOTHELIAL CELLS.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X75593; CAA53266.1; -
DR EMBL; BC000799; AAH00799.1; -
DR HSP; P01112; 1PLL.
DR Genew; HGNC:9762; RAB13.
DR MIM; 602672; -
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR001806; Ras_trnsfrmng.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRfams; TIGR00231; small_GTP; 1.
KW GTP-binding; Lipoprotein; Prenylation; Protein transport;
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 121 124 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 200 200 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 203 AA; 22774 MW; 141621CB998178DA CRC64;

Query Match 48.1%; Score 532; DB 1; Length 203;
Best Local Similarity 47.6%; Pred. No. 4.6e-36;
Matches 100; Conservative 49; Mismatches 43; Indels 18; Gaps 3;

QY 1 MAKQYDVLFRLLLLIGDSGVGKTCCLCRFTDNEFHSHSTIGVDFKMTIEVDGKVRQ 60
Db 1 MAKAYDHLFKLLLLIGDSGVGKTCCLIRFAEDNFNTYISTIGIDFKIRVIEGKKIKLQ 60

QY 61 IWDTAGQERYQTITKYRRAGQIFLVYDISERSYQHIMKWSDVDEYAPGVOKILIG 120
Db 61 VWDTAGQERFKTITTYRRGAMGIIIVYDITDEKSFENIQNMWKSSEKASAGVERLLIG 120

QY 121 NKADEQKRVGREGQOOLAKYGMDFVETSACTNLNINKESFTRTELVLQAHRELEGL 180
Db 121 NKCDMEARKVKQEQADKLAREHGIRFFETSASKSMNVNDEAFSSLDIL-----LKSG 174

QY 181 RMRASNELALAELEEGKPEGPANSSKTC 210
Db 175 GRRSGN-----GNKP--PSTDCLKTC 192

RESULT 5
ID RAO1_DISOM STANDARD; PRT; 200 AA.
AC P22127;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)

DE Ras-related protein ORAL.
OS Discopage ommata (Electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Hynostomalea; Pristiorajae; Batoidea;
OC Torpediniformes; Narcinoidae; Narcinidae; Discopyge.
OX NCBI_TaxID=7785;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Electric lobe;
RX MEDLINE=91115900; PubMed=1899244;
RA Ngsee J.K., Elferink L.A., Scheller R.H.;
RT "A family of ras-like GTP-binding proteins expressed in electromotor
neurons."
RL J. Biol. Chem. 266:2675-2680(1991).
CC -1- SIMILARITY: Ras-RELATED PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M38390; AAA49230.1; -.
DR PIR; A38625; A38625.
DR HSP; P05713; 3RAB.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR SMART; SM00175; RAB; 1.
DR TIGRfams; TIGR00231; small_GTP; 1.
KW GTP-binding; 16 23 GTP (BY SIMILARITY).
FT NP_BIND 16 23 GTP (BY SIMILARITY).
FT NP_BIND 64 68 GTP (BY SIMILARITY).
FT NP_BIND 122 125 GTP (BY SIMILARITY).
FT DOMAIN 122 125 GTP (BY SIMILARITY).
FT DOMAIN 38 46 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 199 199 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 200 200 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 200 AA; 22623 MW; 41D38ED3760519C5 CRC64;

Query Match 48.1%; Score 531.5; DB 1; Length 200;
Best Local Similarity 56.4%; Pred No. 4.9e-36;
Matches 97; Conservative 44; Mismatches 30; Indels 1; Gaps 1;

Qy 1 MAKQ-YDVLFRLLIGDSGVGKTCVLFREFSDDAFNTFTSTIGIDFKIKTVELHGKKIKL 59
Db 1 MAKKTDLFLKLLIGDSGVGKTCVLFREFSDDAFNTFTSTIGIDFKIKTVELHGKKIKL 60

Qy 60 QIWDTAGQERYQTITKQYRRAQIGFLVYDISSERSYQHIMKWSDYDEVAPEGVOKILI 119
Db 61 QIWDTAGQERFHTTTSYRGANGIMLVYDITNAKSFENISKLRNIDEHANEDVERMLL 120

Qy 120 GNKADEKROVREGOQOLAKKEYGMDFYETSACTNINIKESFRLTELVLQ 171
Db 121 GNKCDMDKRVLPKSGEQIAREHAIREFETSANINIEKAFITLAEDILQ 172

RESULT 6
RB10_CANFA STANDARD; PRT; 200 AA.
AC P24409;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Ras-related protein Rab-10.
GN RAB10.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Cocker spaniel;

RX MEDLINE=91061765; PubMed=2123294;
RA Chavrier P., Vingron M., Sander C., Simons K., Zerial M.;
RT "Molecular cloning of YPT1/SEC4-related cDNAs from an epithelial cell
line."
RL Mol. Cell. Biol. 10:6578-6585(1990).
CC -1- FUNCTION: MAY BE INVOLVED IN VESICULAR TRAFFICKING AND
NEUROTRANSMITTER RELEASE.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X56387; CAA39798.1; -.
DR PIR; D36364; D36364.
DR HSP; P05713; 3RAB.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR SMART; SM00175; RAB; 1.
DR TIGRfams; TIGR00231; small_GTP; 1.
KW GTP-binding; Lipoprotein; Prenylation; Protein transport.
FT NP_BIND 16 23 GTP (BY SIMILARITY).
FT NP_BIND 64 68 GTP (BY SIMILARITY).
FT NP_BIND 122 125 GTP (BY SIMILARITY).
FT DOMAIN 38 46 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 199 199 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 200 200 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 200 AA; 22569 MW; 5D52B8E8E47D4362 CRC64;

Query Match 48.0%; Score 530.5; DB 1; Length 200;
Best Local Similarity 55.8%; Pred No. 5.9e-36;
Matches 96; Conservative 46; Mismatches 29; Indels 1; Gaps 1;

Qy 1 MAKQ-YDVLFRLLIGDSGVGKTCVLFREFSDDAFNTFTSTIGIDFKIKTVELHGKKIKL 59
Db 1 MAKKTDLFLKLLIGDSGVGKTCVLFREFSDDAFNTFTSTIGIDFKIKTVELHGKKIKL 60

Qy 60 QIWDTAGQERYQTITKQYRRAQIGFLVYDISSERSYQHIMKWSDYDEVAPEGVOKILI 119
Db 61 QIWDTAGQERFHTTTSYRGANGIMLVYDITNAKSFENISKLRNIDEHANEDVERMLL 120

Qy 120 GNKADEKROVREGOQOLAKKEYGMDFYETSACTNINIKESFRLTELVLQ 171
Db 121 GNKCDMDKRVLPKSGEQIAREHAIREFETSANINIEKAFITLAEDILR 172

RESULT 7
RB10_HUMAN STANDARD; PRT; 200 AA.
AC Q88386; Q9D7X6;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ras-related protein Rab-10.
GN RAB10.
OS Homo sapiens (Human), and
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606; 10090;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Human; TISSUE=Pituitary;
RX MEDLINE=20402571; PubMed=10931946;
RA Hu R.-M., Han Z.-G., Song H.-D., Peng Y.-D., Huang Q.-H., Ren S.-X.,
RA Gu Y.-J., Huang C.-H., Li Y.-B., Jiang C.-L., Fu G., Zhang Q.-H.,


```

DR EMBL; AF035646; AAC29313.1; -.
DR EMBL; AK008725; BAB25858.1; -.
DR HSSP; P05713; 3RAB.
DR Genew; HGNC:9759; RAB10.
DR MGD; MGI:105066; Rab10.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR01806; Ras_trnsfrmng.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRSFRMNG.
DR SMART; SW00175; RAB; 1.
DR TIGRFAMS; TIGR00231; small_gtp; 1.
KW GTP-binding; Lipoprotein; Prenylation; Protein transport.
FT NP_BIND 16 23   GTP (BY SIMILARITY).
FT FT NP_BIND 64 68   GTP (BY SIMILARITY).
FT NP_BIND 122 125   GTP (BY SIMILARITY).
FT DOMAIN 38 46     EFFECTOR REGION (BY SIMILARITY).
FT LIPID 199 199    GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 200 200    GERANYL-GERANYL (BY SIMILARITY).
FT CONFLICT 106 106 N->H (IN REF. 6).
SQ SEQUENCE 200 AA; 22541 MW; 7F02B8E8E4EE1E8 CRC64;

Query Match      48.0%; Score 530.5; DB 1; Length 200;
Best Local Similarity 55.8%; Pred.No.5.e-36;
Matches 96; Conservative 46; Mismatches 29; Indels 1; Gaps

QY 1 MAKQ-YDVLFRLLLGDSGVGKTCCLCRFTDNFEFHSHISTIGVDFKMKTIEVDGIKVRI 59
DB 1 MAKTVDLFELKLIIIGDSGVGKTCVLFERSDDAFNTFTSTIGDKPKITVELQGKKIKL 60
QY 60 QIWDTAGQERYQTITTKQYYRRAGQIFLYVIDISSERSYQHIMKWSVDVEYAPEGVKILI 119
DB 61 QIWDTAGQERFHTITSYYRGAMGLVYDITNGKSFENISKWLNRNIDSHANEDVERMLL 120
QY 120 GNKADEQKRQVGREGOQLAKFYGMDFVETSACTNLNKESPTRLTIELVLQ 171
DB 121 GNKCDDMDKRVPGKEIQIAREHGIRFFETSAKANINIEKAPFLTLAEDILR 172

RESULT 8
RB8B_HUMAN RB8B_HUMAN STANDARD; PRT; 207 AA.
AC Q92930; Q9P293;
DT 15-JUL-1999 (Rel. 38, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ras-related protein Rab-8B.
OS RAB8B.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
[1]
RN RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
CC Seki N., Saito T.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
[2]
RN RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
CC Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
[3]
RN RP SEQUENCE OF 67-119 FROM N.A.
RC TISSUE=Melanoma;
CC Chen D., Guo J., Gahl W.A.;
RL "RAB GTases expressed in human melanoma cells.";
Biochim. Biophys. Acta 1355:1-6(1997).
CC -|- FUNCTION: MAY BE INVOLVED IN VESICULAR TRAFFICKING AND
CC NEUROTRANSMITTER RELEASE (BY SIMILARITY).
-|- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AB038995; BA92249.1; -
 DR EMBL; BC020654; AAH20654.1; -
 DR EMBL; U66624; AAC51199.1; -
 DR HSSP; P05713; 3RAB.
 DR InterPro: IPR003579; GTPase_Rab.
 DR InterPro: IPR001230; Prenyl_site.
 DR InterPro: IPR001806; Ras_trnsmfmg.
 DR InterPro: IPR005225; Small_GTP.
 DR Pfam; PF00071; ras; 1.
 DR PRINTS; PR00449; RASTRNSFRMG.
 DR SMART; SM00175; RAB; 1.
 DR TIGRFAMs; TIGR00231; small_GTP; 1.
 KW GTP-binding; Prenylation; Lipoprotein; Protein transport.
 FT NP_BIND 15 22 GTP (BY SIMILARITY).
 FT NP_BIND 63 67 GTP (BY SIMILARITY).
 FT NP_BIND 121 124 GTP (BY SIMILARITY).
 FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
 FT LIPID 204 204 GERANYL-GERANYL (BY SIMILARITY).
 SQ SEQUENCE 207 AA; 23584 MW; 5960993C0F87F944 CRC64;

Query Match 47.8%; Score 528.5; DB 1; Length 207;
 Best Local Similarity 54.0%; Pred. No. 8.9e-36;
 Matches 95; Conservative 47; Mismatches 33; Indels 1; Gaps 1;

QY 1 MAKQYDVLFRLLIGDSGVKTCCLCRFTDNEPHSHSHISTIGVDFKMKTEVDGKVRQ 60
 DB 1 MAKTYDYLFRLLIGDSGVKTCCLCRFTDNEPHSHSHISTIGVDFKMKTEVDGKVRQ 60
 QY 61 IWDTAGQERYQTITKYRRAAGIFLVYDISSERSYQHIMKWSDVDEYAPEGVQKILIG 120
 DB 61 IWDTAGQERFTITTYAYRGAMGIMLVYDITNEKSFNKNIRNIEEHASSDVERMILG 120
 QY 121 NKADEQKROVGREGQOOLAKEYGMDFYETSACTNLNIKES-FTRLTELVLQAHK 175
 DB 121 NKCDMNDKROVSKERGEKLAIDYGIKLETSKASSNVVEAFFTLARDIMTKLNK 176

RESULT 9
 RB8B_RAT STANDARD; PRT; 207 AA.
 AC P70550;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Ras-related protein Rab-8B.
 GN RB8B.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96393028; PubMed=8799816;
 RA Armstrong J., Thompson N., Squire J.H., Smith J., Hayes B., Solari R.;
 RT "Identification of a novel member of the Rab8 family from the rat
 RT basophilic leukaemia cell line, RBL.2H3.";
 RL J. Cell Sci. 109:1265-1274(1996).
 CC -!- FUNCTION: MAY BE INVOLVED IN VESICULAR TRAFFICKING AND
 CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS OF EXPRESSION IN THE SPLEEN,
 CC TESTIS AND BRAIN.
 CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U53475; AAA99782.1; -
 DR HSSP; P05713; 3RAB.
 DR InterPro: IPR003579; GTPase_Rab.
 DR InterPro: IPR001230; Prenyl_site.
 DR InterPro: IPR001806; Ras_trnsmfmg.
 DR InterPro: IPR005225; Small_GTP.
 DR Pfam; PF00071; ras; 1.
 DR PRINTS; PR00449; RASTRNSFRMG.
 DR SMART; SM00175; RAB; 1.
 DR TIGRFAMs; TIGR00231; small_GTP; 1.
 KW GTP-binding; Prenylation; Lipoprotein; Protein transport.
 FT NP_BIND 15 22 GTP (BY SIMILARITY).
 FT NP_BIND 63 67 GTP (BY SIMILARITY).
 FT NP_BIND 121 124 GTP (BY SIMILARITY).
 FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
 FT LIPID 204 204 GERANYL-GERANYL (BY SIMILARITY).
 SQ SEQUENCE 207 AA; 23603 MW; 4A41AB26BF9DCAF4 CRC64;

Query Match 47.8%; Score 528.5; DB 1; Length 207;
 Best Local Similarity 54.0%; Pred. No. 8.9e-36;
 Matches 95; Conservative 47; Mismatches 33; Indels 1; Gaps 1;

QY 1 MAKQYDVLFRLLIGDSGVKTCCLCRFTDNEPHSHSHISTIGVDFKMKTEVDGKVRQ 60
 DB 1 MAKTYDYLFRLLIGDSGVKTCCLCRFTDNEPHSHSHISTIGVDFKMKTEVDGKVRQ 60
 QY 61 IWDTAGQERYQTITKYRRAAGIFLVYDISSERSYQHIMKWSDVDEYAPEGVQKILIG 120
 DB 61 IWDTAGQERFTITTYAYRGAMGIMLVYDITNEKSFNKNIRNIEEHASSDVERMILG 120
 QY 121 NKADEQKROVGREGQOOLAKEYGMDFYETSACTNLNIKES-FTRLTELVLQAHK 175
 DB 121 NKCDMNDKROVSKERGEKLAIDYGIKLETSKASSNVVEAFFTLARDIMTKLNK 176

RESULT 10
 RB8B_MOUSE STANDARD; PRT; 206 AA.
 AC P55258;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Ras-related protein Rab-8 (Oncogene c-mel).
 GN RB8 OR MEL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91360267; PubMed=1886711;
 RA Nimmo E.R., Sanders P.G., Padua R.A., Hughes D., Williamson R.,
 RA Johnson K.J.;
 RT "The MEL gene: a new member of the RAB/YPT class of RAS-related
 RT genes.";
 RL Oncogene 6:1347-1351(1991).
 CC -!- FUNCTION: MAY BE INVOLVED IN VESICULAR TRAFFICKING AND
 CC NEUROTRANSMITTER RELEASE.
 CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC -----
DR EMBL; S53270; AAB19682.1; -.
DR HSSP; P05713; 3RAB.
DR MGD; MGI:96960; Mel.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR001806; Ras_trnsmfmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMs; TIGR00231; small_gtp; 1.
KW GTP-binding; Prenylation; Lipoprotein; Protein transport;
KW Proto-oncogene.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 121 124 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 203 203 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 206 AA; 23557 MW; 49D832725D662942 CRC64;

Query Match 47.6%; Score 526; DB 1; Length 206;
Best Local Similarity 56.4%; Pred. No. 1.4e-35;
Matches 93; Conservative 41; Mismatches 31; Indels 0; Gaps 0;

QY 1 MAKQYDLVFRLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDFKMTIEVDGIKVRIQ 60
Db 1 MAKTYDLFKLLIGDSGVGKTCVLFRESDAFNSTFTIGIDFKIRTIELDGKRILQ 60

QY 61 IWDTAGQERYQTITKQYRRAGQIFLVYDISSERYOHIMKMWSDVDEYAPGVOKILIG 120
Db 61 IWDTAGQERFRITITAYYRGAMGIMLVYDITNEKSFNIRNIRNIEHASADVEKMILG 120

QY 121 NKADDEQKRVGREGQOOLAKGYGMDVFYETSACTNLNLIKESFTRL 165
Db 121 NKCDVNDKRVQSGEKGLALDYGKFMETSAKANINVENAFTL 165

RESULT 11
YPT2_SCHPO STANDARD; PRT; 200 AA.
AC P17609;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ras-related protein ypt2.
GN YPT2 OR SPAC9B9.07C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=90269232; PubMed=2112089;
RA Hengst L., Lehmeier T., Gallwitz D.;
RT "Structural and functional analysis of ypt2, an essential ras-related
RT gene in the fission yeast Schizosaccharomyces pombe encoding a Sec4
RT protein homologue.";
RL EMBO J. 9:1957-1962(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90332438; PubMed=2115995;
RA Fawell E., Hook S., Sweet D., Armstrong J.;
RT "Novel Ypt1-related genes from Schizosaccharomyces pombe.";
RL Nucleic Acids Res. 18:4264-4264(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

```

```

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., Mclean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Calibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
CC TRAFFIC (BY SIMILARITY).
CC -!- MISCELLANEOUS: THIS PROTEIN IS ESSENTIAL FOR CELL VIABILITY.
CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC PROBABLE YEAST SEC4 HOMOLOG.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52469; CAA36707.1; -.
DR EMBL; X52864; CAA37045.1; -.
DR EMBL; Z99262; CAB16405.1; -.
DR PIR; S10493; S10493.
DR PIR; S12790; S12790.
DR HSSP; P05713; 3RAB.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsmfmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMs; TIGR00231; small_gtp; 1.
KW GTP-binding; Lipoprotein; Prenylation; Protein transport.
FT NP_BIND 16 23 GTP (BY SIMILARITY).
FT NP_BIND 64 68 GTP (BY SIMILARITY).
FT NP_BIND 122 125 GTP (BY SIMILARITY).
FT DOMAIN 38 46 EFFECTOR REGION (PROBABLE).
FT LIPID 199 199 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 200 200 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 200 AA; 22755 MW; 2C658D153A290C30 CRC64;

Query Match 46.2%; Score 510; DB 1; Length 200;
Best Local Similarity 51.9%; Pred. No. 2.6e-34;
Matches 97; Conservative 39; Mismatches 43; Indels 8; Gaps 1;

QY 3 KOYDVLFRLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDFKMTIEVDGIKVRIQ 62
Db 3 KOYDVLFRLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDFKMTIEVDGIKVRIQ 62
QY 4 KSYDYLKLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDFKMTIEVDGIKVRIQ 63
Db 4 KSYDYLKLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDFKMTIEVDGIKVRIQ 63
QY 63 DFAGQERYQTITKQYRRAGQIFLVYDISSERYOHIMKMWSDVDEYAPGVOKILIGNK 122
Db 63 DFAGQERYQTITKQYRRAGQIFLVYDISSERYOHIMKMWSDVDEYAPGVOKILIGNK 122
QY 64 DFAGQERYQTITKQYRRAGQIFLVYDISSERYOHIMKMWSDVDEYAPGVOKILIGNK 123
Db 64 DFAGQERYQTITKQYRRAGQIFLVYDISSERYOHIMKMWSDVDEYAPGVOKILIGNK 123
QY 123 ADEEQKRVGREGQOOLAKGYGMDVFYETSACTNLNLIKESFTRLTLVLQAHKLEGLRM 182

```



```

DR PIR; JQ0908; JQ0908.
DR PIR; JS0640; JS0640.
DR HSSP; P05713; 3RAB.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsfmrng.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMS; TIGR00231; small_GTP; 1.
KW GTP-binding; Lipoprotein; Prenylation; Multigene family.
FT NP_BIND 22 29 GTP (BY SIMILARITY).
FT NP_BIND 70 74 GTP (BY SIMILARITY).
FT NP_BIND 128 131 GTP (BY SIMILARITY).
FT DOMAIN 44 52 EFFECTOR REGION (BY SIMILARITY).
FT DOMAIN 213 213 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 214 214 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 216 216 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 216 AA; 23935 MW; 09E9C19A7A44E705 CRC64;

Query Match 46.0%; Score 508; DB 1; Length 216;
Best Local Similarity 46.9%; Pred. No. 4.2e-34;
Matches 100; Conservative 41; Mismatches 54; Indels 18; Gaps 3;

QY 5 YDVLFRLLIGDSGVGKCTLLCRFTDNEFHSHSHISTIGVDFKMKTIYVDGKIVRIQIWD 64
DB 12 YDYLKLLIGDSGVGKSCLLRFRSDGFTTSFTITIGIDFKIRTIYELDKRIKLIQIWD 71
QY 65 AGQERYQITTKYYRRAQGIIFLVYDISSERSYOHIMKWSDVDYAPGVQVKILIGNKAD 124
DB 72 ACQERFRITTYAYRGAGIILVYDITDESSFNIRNIRNIEQHASDNVNKILVGNKAD 131
QY 125 -BEQRQVREGQOQLAKEYGMDFYETSACTNLNKESTFRTLTELVLQAKRKEGLELRMR 183
DB 132 MDESRAVPTAKGQALADEYGIKFFETSAKTNLNVVEEVSIG-----RDIQR 180
QY 184 ASNELALAE-----LEEEGKPEGPANSSKTC 210
DB 181 LSDTDSRAEPATIKISOTDAQAGAQATQKSAC 213

RESULT 14
YPT2_VOLCA
ID YPT2_VOLCA STANDARD; PRT; 217 AA.
AC P36861;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE GTP-binding protein yptv2.
GN YPTV2.
OS Volvox carteri.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Volvocaceae; Volvox.
OX NCBI_TaxID=3067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=f. Nagariensis / HK10;
RX MEDLINE=94037148; PubMed=8221932;
RA Fabry S., Jacobsen A., Huber H., Palme K., Schmitt R.;
RT "Structure, expression, and phylogenetic relationships of a family of
RT ypt genes encoding small G-proteins in the green alga Volvox
RT carteri.";
RL Curr. Genet. 24:229-240(1993).
CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
CC -!- TRAFFIC (BY SIMILARITY).
CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; L08128; AAA34251.1; .
DR PIR; S36365; S36365.
DR HSSP; P05713; 3RAB.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsfmrng.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMS; TIGR00231; small_GTP; 1.
KW GTP-binding; Lipoprotein; Prenylation; Protein transport;
Multigene family.
FT NP_BIND 20 27 GTP (BY SIMILARITY).
FT NP_BIND 68 72 GTP (BY SIMILARITY).
FT NP_BIND 126 129 GTP (BY SIMILARITY).
FT DOMAIN 42 50 EFFECTOR REGION (PROBABLE).
FT LIPID 215 215 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 216 216 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 217 AA; 24143 MW; 87D3B30C75689EAA CRC64;

Query Match 45.7%; Score 505.5; DB 1; Length 217;
Best Local Similarity 48.6%; Pred. No. 6.7e-34;
Matches 101; Conservative 40; Mismatches 64; Indels 3; Gaps 3;

QY 5 YDVLFRLLIGDSGVGKCTLLCRFTDNEFHSHSHISTIGVDFKMKTIYVDGKIVRIQIWD 64
DB 10 YDALIKLLVSGVGKSCLLRFRSDGFTTSFTITIGIDFKIKKVDYDGLVKLIQIWD 69
QY 65 AGQERYQITTKYYRRAQGIIFLVYDISSERSYOHIMKWSDVDYAPGVQVKILIGNKAD 124
DB 70 ACQERFRITTYAYRGAGIILVYDITDESSFNIRNIRNIEQHASDNVNKILVGNKLD 129
QY 125 -BEQRQVREGQOQLAKEYGMDFYETSACTNLNKESTFRTLTELVLQAKRKEGLELRMR 183
DB 130 LAEDKRVYSIARGQALADEYGFYETSAKDNVHVEEAFIAVAKDVLARMEGEHANQQL 189
QY 184 ASNELALAE-LEEEGKPEGPANSSKTC 210
DB 190 QOOQLSAAQVRLTSGSP-SPAQGSKC 216

RESULT 15
RE10_RAT
ID RB10_RAT STANDARD; PRT; 200 AA.
AC P35281;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Ras-related protein Rab-10.
GN RAB10.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; Tissue=Brain;
RX MEDLINE=92210533; PubMed=1313420;
RA Elferink L.A., Anzai K., Scheller R.H.;
RT "Rab15, a novel low molecular weight GTP-binding protein specifically
RT expressed in rat brain.";
RL J. Biol. Chem. 267:5768-5775(1992).
CC -!- FUNCTION: MAY BE INVOLVED IN VESICULAR TRAFFICKING AND
CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS IN NEURAL AND MUSCLE TISSUES.
CC -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_email_to_license@isb-sib.ch).

```

-----
DR EMBL; M83677; AAA41991.1; -.
DR PIR; B42148; B42148.
DR HSP; P05713; 3RAB.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
KW GTP-binding; Lipoprotein; Prenylation; Protein transport.
FT NP_BIND 16 23 GTP (BY SIMILARITY).
FT NP_BIND 54 68 GTP (BY SIMILARITY).
FT NP_BIND 122 125 GTP (BY SIMILARITY).
FT DOMAIN 38 46 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 199 199 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 200 200 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 200 AA; 22858 MW; EFAP804FD4C3B6A3 CRC64;

Query Match 45.5%; Score 502.5; DB 1; Length 200;
Best Local Similarity 53.5%; Pred. No. 1.1e-33;
Matches 92; Conservative 47; Mismatches 32; Indels 1; Gaps 1;

QY 1 MAKQ-YDVLFRLLIGDSGVGKTCILCRFTDNEFHSHSTIGVDPKMTIEVDGIKVR I 59
DB 1 MAKTYDLLFKLLIGDSGVGKTCVLFREFSDAFNTFTIEIDFKITVELQGGKIKL 60

QY 60 QIWDTAGQERYQITKQYVRAOGIFLVYDISSERYQHIMKWSDVDEYAPEGVOKILI 119
DB 61 QIWDTAGQERFHTITTSYIRGANGIMLYVDITNGKSFENISKWLNRIDQHANEDVERMLL 120

QY 120 GNKADEEQKRGVREGOQOLAKKEYGMDFYETSACTNLNIKESFTRLTELVLQ 171
DB 121 RNKCDMDHKRVKPKGQEIAREHRIREFETSAKANINIEKAPLTPEDILR 172

```

Search completed: November 17, 2002, 20:39:12
Job time : 26 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model
Run On: November 17, 2002, 20:35:44 ; Search time 43 Seconds
(without alignments)
473.965 Million cell updates/sec

Title: US-09-817-198A-2
Perfect score: 1105
Sequence: 1 MAKQYDVLFRLLIGDSGVG.....LEEKGKPGPANSSKTCWC 212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1077	97.5	212	2 F42148	GTP-binding protei
2	559	50.6	224	2 T33855	hypothetical prote
3	545	49.3	207	2 B49647	GTP-binding protei
4	545	49.3	207	2 B36364	GTP-binding protei
5	532	48.1	203	2 A49647	GTP-binding protei
6	531.5	48.1	200	2 A38625	GTP-binding protei
7	530.5	48.0	200	2 D36364	GTP-binding protei
8	530	48.0	209	2 B38625	GTP-binding protei
9	526	47.6	206	2 I78851	GTP-binding protei
10	522.5	47.3	200	2 B42148	GTP-binding protei
11	519	47.0	216	2 T45901	GTP-binding protei
12	514	46.5	201	2 T28971	hypothetical prote
13	512	46.3	216	2 S33900	GTP-binding protei
14	512	46.3	216	2 T48378	GTP-binding protei
15	510	46.2	200	2 S12790	GTP-binding protei
16	509	46.1	215	2 T14565	GTP-binding protei
17	508.5	46.0	215	2 S7478	GTP-binding protei
18	508	46.0	216	2 J50640	GTP-binding protei
19	507.5	45.9	222	2 T14405	small GTP-binding
20	506	45.8	216	2 S57471	GTP-binding protei
21	505.5	45.7	217	2 S36365	GTP-binding protei
22	502.5	45.5	215	2 S7462	GTP-binding protei
23	501	45.3	203	2 S1495	GTP-binding protei
24	495.5	44.8	204	2 J7589	Sec4p homolog - ye
25	492.5	44.6	215	2 S57474	GTP-binding protei
26	483.5	43.8	208	2 A34716	GTP-binding protei
27	483	43.7	208	2 A38202	GTP-binding protei
28	482	43.6	203	2 S34253	GTP-binding protei
29	480	43.4	202	2 S38740	GTP-binding protei

30	479	43.3	203	2 B34716	GTP-binding protei
31	478	43.3	202	2 S72515	GTP-binding protei
32	477.5	43.2	201	2 JC2488	GTP-binding protei
33	477.5	43.2	203	2 JC4105	GTP-binding protei
34	477	43.2	201	2 JC5337	GTP-binding protei
35	477	43.2	202	2 S41430	GTP-binding protei
36	476	43.1	258	2 B86153	ARA-5 [imported] -
37	475.5	43.0	203	2 JC1247	GTP-binding protei
38	475	43.0	203	2 S30096	GTP-binding protei
39	473.5	42.9	205	2 T33781	hypothetical protei
40	473	42.8	206	2 T14391	GTP-binding protei
41	473	42.8	210	2 T18242	ras protein homolo
42	470.5	42.6	203	2 B38202	GTP-binding protei
43	470.5	42.6	215	1 TVBYQ4	GTP-binding protei
44	469	42.4	205	2 S38339	GTP-binding protei
45	468	42.4	205	1 TVRTYP	GTP-binding protei

ALIGNMENTS

RESULT 1

F42148

GTP-binding protein rab15 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 02-Feb-2001

C:Accession: F42148

R:Elferink, L.A.; Anzai, K.; Scheller, R.H.

J. Biol. Chem. 267, 5768-5775, 1992

A:Title: rab15, a novel low molecular weight GTP-binding protein specifically express

A:Reference number: A42148; MUID:92210533; PMID:1313420

A:Accession: F42148

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-212 <ELF>

C:Cross-references: GB:M83679; NID:g206536; PIDN:AAA41995.1; PID:g206537

C:Superfamily: ras transforming protein; translation elongation factor Tu homology

C:Keywords: GTP binding; lipoprotein; nucleotide binding; P-loop; prenylated cysteine

F:9-124/Domain: translation elongation factor Tu homology <ETU>

F:15-22/Region: nucleotide-binding motif A (P-loop)

F:121-124/Region: GTP-binding NXKD motif

F:210,212/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 97.5%; Score 1077; DB 2; Length 212;
Best Local Similarity 97.6%; Pred. No. 9.2e-79;
Matches 207; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAKQYDVLFRLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDFKMKKTIEVDGIKVRIQ 60

Db 1 MAKQYDVLFRLLIGDSGVGKTCLLCRFTDNEFHSSHISTIGVDFKMKKTIEVDGIKVRIQ 60

QY 61 IWDTAGORYQITQYVYRAQGIYLVYDISSERSYQHIMKWSDVDEYAPGVQKILIG 120

Db 61 IWDTAGORYQITQYVYRAQGIYLVYDISSERSYQHIMKWSDVDEYAPGVQKILIG 120

QY 121 NKADEQKRVGREGQQQLAKGYMDFYETSACTNLNLIKESFTRLTETVLQAHKREGL 180

Db 121 NKADEQKRVGREGQQQLAKGYMDFYETSACTNLNLIKESFTRLTETVLQAHKREGL 180

QY 181 RMRASNELALAELEEEGKPGPANSSKTCWC 212

Db 181 RTCASNELALAELEDEGKTEGPANSSKTCWC 212

RESULT 2

T33855

hypothetical protein D1037.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 11-Jan-2000

C:Accession: T33855

R:Liedwith, J.; Biewald, T.

submitted to the EMBL Data Library, November 1998

A:Description: The sequence of C. elegans cosmid D1037.

A:Reference number: 221424
A:Accession: T33855
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-224 <LED>
A:CROSS-references: EMBL:AF106592; PIDN:AAC78494.1; GSPDB:GN00019; CESP:D1037.4
A:Experimental source: strain Bristol N2; clone D1037
C:Genetics:
A:Gene: CESP:D1037.4
A:Map position: 1
A:Introns: 10/3; 62/2; 82/3; 181/1
C:Superfamily: ras transforming protein; translation elongation factor Tu homology

Query Match 50.6%; Score 559; DB 2; Length 224;
Best Local Similarity 55.3%; Pred. No. 2.1e-37;
Matches 105; Conservative 42; Mismatches 39; Indels 4; Gaps 1;

QY 1 MAKQYDLVRLLLIGDSGVGKTCCLCRFTDNEFHSHHISTIGVDFKMKITIEVDGKIKVRIQ 60
DB 1 MAKTYDYLKLLIGDSGVGKTCVLFREFSDSFNNSTIGIDFKIRTIELDGKKIKLQ 60
QY 61 IWDTAGQERYQTITKQYRRAGIFLYVDISSERSYQHIMKWSDVDEYAPGVOKILIG 120
DB 61 IWDTAGQERFTITTTAYYRGAMGIIIVDITNERSFENIKNIRNIEEHAASDVEMKILG 120
QY 121 NKADEQKROVGRGQOQLAKEYGMDFYETSACTNINIKESFTRLTFLVLAQHRKELEGL 180
DB 121 NKCDIERREVSRRGEQLAIEYGTFLFETSAKANLNIDEAFPTLARDI----KSKMEQN 176
QY 181 RMRASNELAL 190
DB 177 EMRAATGAAI 186

RESULT 3

B49647
GTP-binding protein rab8 - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 19-Jan-2001
C:Accession: B49647; S36817
R:Zahraoui, A.; Joberty, G.; Arpin, M.; Fontaine, J.J.; Hellio, R.; Tavitian, A.; Louvar
J. Cell Biol. 124, 101-115, 1994
A:Title: A small rab GTPase is distributed in cytoplasmic vesicles in non polarized cell
A:Reference number: A49647; MUID:94124602; PMID:8294494
A:Accession: B49647
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-207 <ZAH>
A:CROSS-references: EMBL:X56741; NID:g452317; PIDN:CAA40065.1; PID:g452318
R:Joberty, G.; Tavitian, A.; Zahraoui, A.
FEBS Lett. 330, 323-328, 1993
A:Title: Isoprenylation of Rab proteins possessing a C-terminal Caax motif.
A:Reference number: S36817; MUID:93387463; PMID:8375503
A:Accession: S36817
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 175-186 <JOB>
C:Superfamily: ras transforming protein; translation elongation factor Tu homology
C:Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop;
F:9-124/Domain: translation elongation factor Tu homology <ETU>
F:15-22/Region: nucleotide-binding motif A (P-loop)
F:121-124/Region: GTP-binding NKXD motif
F:151-153/Region: GTP-binding SAK/L motif
F:204/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 49.3%; Score 545; DB 2; Length 207;
Best Local Similarity 52.9%; Pred. No. 2.4e-36;
Matches 99; Conservative 45; Mismatches 43; Indels 0; Gaps 0;

QY 1 MAKQYDLVRLLLIGDSGVGKTCCLCRFTDNEFHSHHISTIGVDFKMKITIEVDGKIKVRIQ 60
DB 1 MAKTYDYLKLLIGDSGVGKTCVLFREFSDAFNSTFTSTIGIDFKIRTIELDGKKIKLQ 60

QY 61 IWDTAGQERYQTITKQYRRAGIFLYVDISSERSYQHIMKWSDVDEYAPGVOKILIG 120
DB 61 IWDTAGQERFTITTTAYYRGAMGIIIVDITNERSFENIKNIRNIEEHAASDVEMKILG 120
QY 121 NKADEQKROVGRGQOQLAKEYGMDFYETSACTNINIKESFTRLTFLVLAQHRKELEGL 180
DB 121 NKCDVNDKROVSKERGEKALDYGKFMETSAKANINVENAFPTLARDIKAKMDKKLEGN 180
QY 181 RMRASNE 187
DB 181 SPQGSNQ 187
RESULT 4
B36364
GTP-binding protein rab8 - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 02-Feb-2001
C:Accession: B36364; S15604
R:Chavrier, P.; Vingron, M.; Sander, C.; Simons, K.; Zerial, M.
Mol. Cell. Biol. 10, 6578-6585, 1990
A:Title: Molecular cloning of YPT1/SEC4-related cDNAs from an epithelial cell line.
A:Reference number: A36364; MUID:91061765; PMID:2123294
A:Accession: B36364
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-207 <CHA>
A:CROSS-references: GB:X56385; NID:9920; PIDN:CAB56776.1; PID:g6006436
C:Superfamily: ras transforming protein; translation elongation factor Tu homology
C:Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop
F:9-124/Domain: translation elongation factor Tu homology <ETU>
F:15-22/Region: nucleotide-binding motif A (P-loop)
F:121-124/Region: GTP-binding NKXD motif
F:151-153/Region: GTP-binding SAK/L motif
F:204/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 49.3%; Score 545; DB 2; Length 207;

Best Local Similarity 52.9%; Pred. No. 2.4e-36;
Matches 99; Conservative 45; Mismatches 43; Indels 0; Gaps 0;

QY 1 MAKQYDLVRLLLIGDSGVGKTCCLCRFTDNEFHSHHISTIGVDFKMKITIEVDGKIKVRIQ 60
DB 1 MAKTYDYLKLLIGDSGVGKTCVLFREFSDAFNSTFTSTIGIDFKIRTIELDGKKIKLQ 60
QY 61 IWDTAGQERYQTITKQYRRAGIFLYVDISSERSYQHIMKWSDVDEYAPGVOKILIG 120
DB 61 IWDTAGQERFTITTTAYYRGAMGIIIVDITNERSFENIKNIRNIEEHAASDVEMKILG 120
QY 121 NKADEQKROVGRGQOQLAKEYGMDFYETSACTNINIKESFTRLTFLVLAQHRKELEGL 180
DB 121 NKCDVNDKROVSKERGEKALDYGKFMETSAKANINVENAFPTLARDIKAKMDKKLEGN 180
QY 181 RMRASNE 187
DB 181 SPQGSNQ 187

RESULT 5

A49647
GTP-binding protein Rab13 - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 02-Feb-2001
C:Accession: A49647
R:Zahraoui, A.; Joberty, G.; Arpin, M.; Fontaine, J.J.; Hellio, R.; Tavitian, A.; Lou
J. Cell Biol. 124, 101-115, 1994
A:Title: A small rab GTPase is distributed in cytoplasmic vesicles in non polarized c
A:Reference number: A49647; MUID:94124602; PMID:8294494
A:Accession: A49647
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-203 <ZAH>
A:CROSS-references: EMBL:X75593; NID:g452319; PIDN:CAA53266.1; PID:g452320
C:Superfamily: ras transforming protein; translation elongation factor Tu homology

Query Match	46.3%	Score 512;	DB 2;	Length 216;
Best Local Similarity	49.8%	Pred. No. 1.le-33;		
Matches 100;	Conservative 43;	Mismatches 46;	Indels 12;	Gaps 3;

Qy	65	AGBRYQTITKQYRRAGCFLVYDISSRSYQHIMTKWVSVDVEYAPGVQKILIGNKAD	124
		:	
		:	
		:	
		:	
		:	
Db	72	AGRFRFRITITAYRGAMGILLVYDTDESFNNIRNMKNKIEGHASDNVNKILVGNKAD	131
		:	
		:	
		:	
		:	
		:	

QY 176 ELEGRLMRASNELALAELEEE 196

QY 176 ELEGLRMRASNELALAELEEE 196
| : | : : | : | :
Db 189 EPOGIKITKODTAASSSTAEEK 209

RESULT 15
S12790 GTP-binding protein ypt2 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 02-Feb-2001
C:Accession: S12790; S10493; T39214
R:Haubruck, H.; Engelke, U.; Mertins, P.; Gallwitz, D.
EMBO J. 9, 1957-1962, 1990

EMBO J. 9, 1937-1944, 1990
A>Title: Structural and functional analysis of ypt2, an essential ras-
A;Reference number: S12790; MUID:90269232; PMID:2112089

A, molecule type: DNA
A;Residues: 1-200 <HAU>
A;Cross-references: EMBL:X52469; NID:g5143; PID:CAA36707.1; PID:g5144
R;Fawell, E.; Hook, S.; Sweet, D.; Armstrong, J.
Nucleic Acids Res. 18, 4264, 1990

A; Title: Novel YPT1-related genes from Schizosaccharomyces pombe.
Nucleic Acids Res. 18, 4204, 1990
A; Reference number: S10492; MUID:90332438; PMID:2115995

A;Accession: S10493
A;Molecule type: DNA

A:Residues: 1-200 <FAW>
A:Cross-references: EMBL:X52864; NID:g5149; PIDN:CAA37045.1; PID:g5150
B:McDougal P. , Barrell B. C. , Batandream M. A. , Wood V.

K, McDougal, K.; Bailett, B.S.; Rajandream, R.S. submitted to the EMBL Data Library, September 1997
A: Reference number: Z21836

A;Accession: T39214
A;Status: preliminary; tra

A;Molecule type: DNA

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2002, 20:42:04 ; Search time 51 seconds
(without alignments)
1536.011 Million cell updates/sec

Title: US-09-817-198A-2
Perfect score: 1105
Sequence: 1 MAKQYDVLFRLLIGDSGVG.....LEEKGKPEGPANSSKTCWC 212

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 328367 seqs, 184756068 residues

Total number of hits satisfying chosen parameters: 656734

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPRO_spool/US09817198/runat_13112002_135207_7068/app_query.fasta_1.391
-DB=PublishedApplications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09817198@cgn.1.1.20 @runat_13112002_135207_7068
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA: *
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq: *
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq: *
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq: *
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq: *
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq: *
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq: *
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1105	100.0	3257	10	US-09-817-198A-1
2	534.5	48.4	1274	10	US-09-925-302-91
3	530.5	48.0	1537	10	US-09-925-300-631
4	528.5	47.8	624	10	US-09-794-257-9

5	528.5	47.8	1161	10	US-09-794-257-7	Sequence 7, Appli
6	528.5	47.8	2497	10	US-09-834-975-879	Sequence 879, App
7	528.5	47.8	2497	10	US-09-834-975-885	Sequence 885, App
8	528.5	47.8	2497	10	US-09-834-975-894	Sequence 894, App
9	528.5	47.8	2497	10	US-09-834-975-896	Sequence 896, App
10	519	47.0	651	9	US-09-938-842A-836	Sequence 836, App
11	478.5	43.3	925	10	US-09-967-736-4	Sequence 4, Appli
12	477.5	43.2	881	10	US-09-770-445-529	Sequence 529, App
13	468	42.4	609	9	US-09-938-842A-832	Sequence 832, App
14	448	40.5	585	10	US-09-917-800A-1461	Sequence 1461, Ap
15	429	38.8	875	12	US-10-051-986-10	Sequence 10, Appl
16	429	38.8	1116	10	US-09-794-257-13	Sequence 13, Appl
17	429	38.8	2674	10	US-09-817-199A-1	Sequence 1, Appli
18	419	37.9	639	10	US-09-350-874-66	Sequence 66, Appl
19	416	37.6	576	10	US-09-794-257-15	Sequence 15, Appl
20	415.5	37.6	894	10	US-09-770-445-487	Sequence 487, App
21	412.5	37.3	538	10	US-09-924-035A-794	Sequence 794, App
22	411.5	37.2	654	9	US-09-938-842A-2113	Sequence 2113, Ap
23	406	36.7	896	10	US-09-770-445-478	Sequence 478, App
24	404.5	36.6	3936	10	US-09-919-172-49	Sequence 49, Appl
25	399	36.1	3124	10	US-09-925-302-340	Sequence 340, App
26	395	35.7	844	10	US-09-770-445-646	Sequence 646, App
27	391	35.4	1673	9	US-09-954-531-1359	Sequence 1359, Ap
28	388.5	35.2	4083	10	US-09-817-182-1	Sequence 1, Appli
29	387.5	35.1	654	9	US-09-938-842A-78	Sequence 78, Appl
30	387	35.0	771	10	US-09-728-445-652	Sequence 652, App
31	384.5	34.8	1042	10	US-09-954-456-563	Sequence 563, App
32	382.5	34.6	676	10	US-09-770-149-374	Sequence 374, App
33	381	34.5	645	9	US-09-938-842A-774	Sequence 774, App
34	381	34.5	900	10	US-09-770-445-469	Sequence 469, App
35	379	34.3	577	10	US-09-998-1598-1231	Sequence 1231, Ap
36	378.5	34.3	857	10	US-09-917-800A-1426	Sequence 1426, Ap
37	376	34.0	642	10	US-09-794-257-6	Sequence 6, Appli
38	376	34.0	1023	10	US-09-794-257-4	Sequence 4, Appli
39	374	33.8	1022	10	US-09-920-300A-1703	Sequence 1703, Ap
40	374	33.8	1022	12	US-10-033-528-1703	Sequence 1703, Ap
41	373	33.8	847	10	US-09-988-974-4	Sequence 4, Appli
42	373	33.8	1129	10	US-09-925-301-235	Sequence 235, App
43	372	33.7	660	9	US-09-938-842A-901	Sequence 901, App
44	371	33.6	1061	10	US-09-880-107-3393	Sequence 3393, Ap
45	370	33.5	2721	10	US-09-822-860-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-817-198A-1
; Sequence 1, Application US/09817198A
; Patent No. US20020146758A1
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; FILE REFERENCE: CLO01188
; CURRENT APPLICATION NUMBER: US/09/817,198A
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3257
; TYPE: DNA
; ORGANISM: Human
US-09-817-198A-1

Alignment Scores:
Pred. No.: 2.14e-129 Length: 3257
Score: 1105.00 Matches: 212
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0


```
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 1046
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 896
; LENGTH: 2497
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2497)
; OTHER INFORMATION: n = A,T,C or G
US-09-834-975-896

Alignment Scores:
Pred. No.: 6,62e-57 Length: 2497
Score: 528.50 Matches: 95
Percent Similarity: 80.68% Conservative: 47
Best Local Similarity: 53.98% Mismatches: 33
Query Match: 47.83% Indels: 1
DB: 10 Gaps: 1

US-09-817-198A-2 (1-212) x US-09-834-975-896 (1-2497)
QY 1 MetAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerGlyValGlyThrCysLeu 20
Db 92 ATGCGAAGACGACGATTAATCTCTTCAAGCTCTCTGATCGCGACTCGGGGGTAGGC 151
QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisIleSerThr 40
Db 152 AAGACCTGCCTCTCTGTTCCCTCTCAGAGGACGCGCTTCAACACACCTTCATCCACC 211
QY 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGln 60
Db 212 ATCCGAATGATTTAAATATAGACATAGACTAGTGAAGAAATAAGCTTCAG 271
QY 61 IleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
Db 272 ATATGGGACACAGCGGTCAGGAAGATTCGAAACATCACGACGCGTACTACAGAGGA 331
QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
Db 332 GCCATGGGCATATGCTGGTCTATGACATCACAATGAAAATCCCTTTGGACATATATAA 391
QY 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
Db 392 AATTGGATCAGAAACATTTGAAGACATGCTCTCCGATGTCGAAAGAAATGATCCTGGGT 451
QY 121 AsnLysAlaAspGluGlnGlnLysArgGlnValGlyArgGlnGlnGlnGlnLeuAla 140
Db 452 AACAAATGTGATATGAATGACAAAGACAAAGTGTCAAAGAAAGAGGGGAGAGCTAGCA 511
QY 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
Db 512 ATTGACTATGGGATTAATCTTGGAGACAAAGCGCAAAATCCATCGCAATGTAGAGAG 571
QY 161 Ser---PheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLys 175
Db 572 GCATTTTATACATTGCGACGAGATATATGACAAAACTCAACAGAAAA 619

RESULT 10
US-09-938-842A-836
; Sequence 836, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
```

```
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 836
; LENGTH: 651
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-836

Alignment Scores:
Pred. No.: 1.51e-56 Length: 651
Score: 519.00 Matches: 102
Percent Similarity: 71.64% Conservative: 42
Best Local Similarity: 50.75% Mismatches: 42
Query Match: 46.97% Indels: 15
DB: 9 Gaps: 3

US-09-817-198A-2 (1-212) x US-09-938-842A-836 (1-651)
QY 5 TyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerGlyValGlyLysThrCysLeu 24
Db 34 TAGCATTAACCTATTAAACTTCTGCTGATCGGAGACGCGGTGTGGTAAGATTGCCCTT 93
QY 25 LeuCysArgPheThrAspAsnGluPheHisSerHisIleSerThrIleGlyValAsp 44
Db 94 CTTCTACGTTCTCAGATGGCTGTTTACCACCACTTTTATTACCACTATTGGGATTGAT 153
QY 45 PheLysMetLysThrIleGluValAspGlyIleLysValArgIleGlnIleTrpAspThr 64
Db 154 TTTAAGATACCGCATATTGAGCTCGATGGGAAGAAATTAAGCTGCAATCTGGGATACT 213
QY 65 AlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgAlaGlnGlyIle 84
Db 214 GCGGACAGAGCGGTTCGCGACATCACAACTGCTACTACCGTGGAGCCATGGGATT 273
QY 85 PheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMetLysTrpValSer 104
Db 274 TTGCTTGTGTATGATGTGATGATGAATCATCTTTCAACAACATCAGGAATTTGGATCCGT 333
QY 105 AspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGlyAsnLysAlaAsp 124
Db 334 AACATTGAGCAGATGCTTCTGTATAGTGTCAACAAGATTTCTAGTTGGGAACAACAGAT 393
QY 125 ---GluGluGlnLysArgGlnValGlyArgGlnGlnGlnGlnGlnLeuAlaLysGluTyr 143
Db 394 ATGGATGAAGCAAAAGAGCGTGTGCCAAAATCTAAGGGCCCAAGCTCTTGCAGATGAATAT 453
QY 144 GlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGluSerPhe-Th 163
Db 454 GGAATGAAGTTTTCGAGACTAGTGCACAAGCTTAACCTTAACCTTGAGGAAGTTTCTTTC 513
QY 163 rArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGlyLeuArgMetAr 183
Db 514 T-----CTATGTAAAGACATTAACCAAGACTTG----- 544
QY 183 gAlaSerAsnGluLeuAlaLeuAlaGluLeuGluGluGluGluGluGluGluGluGluGlu 203
Db 545 -----CAGATACCGATGCGAGCTGAGCGCGCAAAACAATCAAAAATCAACCAATCC 594
QY 203 o 203
Db 595 G 595

RESULT 11
US-09-967-736-4
; Sequence 4, Application US/09967736
; Patent No. US20020103340A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
```

Lal, Preeti
Corley, Neil C.
Shah, Purvi
TITLE OF INVENTION: RAB PROTEINS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
FILING DATE: 28-Sep-2001
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 09/154,602
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0367 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-853-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 925 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LIVRUT04
CLONE: 2514506
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-967-736-4

Alignment Scores:
Pred. No.: 3,18e-51 Length: 925
Score: 478.50 Matches: 91
Percent Similarity: 65.84% Conservative: 42
Best Local Similarity: 45.09% Mismatches: 64
Query Match: 43.30% Indels: 5
DB: 10 Gaps: 1

US-09-817-198A-2 (1-212) x US-09-967-736-4 (1-925)

QY 1 MetAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuGlyAspSerGlyValGly 20
Db 66 ATGAACCCGATATGACTACCTGTTTAACTGCTTTTATGGCGACTCAGCGCTGGGC 125
QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisLeuSerThr 40
Db 126 AAGTCATGCTGCTCTCGGGTTTCTGATGACAGCTACAGAGAGCTACATCAGCACC 185
QY 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGln 60
Db 186 ATCGGGTGGACTTCAGATCCGACCATCGAGCTGGGCAAACTATCAACTTCAG 245
QY 61 IleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
Db 246 ATCTGGACACAGCGCGCCAGGAAGGTTCGCGACCATCACTTCCAGCTACTACCGGGG 305
QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
Db 306 GCTATGGCATCATCTGGTGTATACGCTACAGGATCATCTACGCCAACCTGAAG 365
QY 101 LysTrpValSerAspValAspGlnTyrAlaProGluGlyValGlnLysIleLeuIleGly 120

Db 366 CAGTGGCTGCAGGAGATTGACCGCTATGCCAGCAGAGACGTCAATAAGCTCTGGTGGC 425
QY 121 AsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlyGlnGlnLeuAla 140
Db 426 AACAAAGAGCGACCTCACCACCAAGAGGTGGACAAACACACACCAAGAGTTTGA 485
QY 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
Db 486 GACTCTCTGGGCATCCCTTCTTGGAGAGGAGGCAAGAAATGCCACCAATGTCGAGCAG 545
QY 161 SerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGlnLeuGlyLeu 180
Db 546 GCGTTC-----ATGACCATGCTGCTGAAATCAAAAGCGGATGGGCGCT 590
QY 181 ArgMetArgAlaSerAsnGluLeuAlaLeuAlaGluLeuGluGluGluGlyLysPro 200
Db 591 GGAGCAGCCTCTGGGGGCGAGCGGCCCAATCTCAAGATGCGACGACGCCCTGTAAAGCG 650
QY 201 GluGly 202
Db 651 GCTGGC 656

RESULT 12

US-09-770-445-529
; Sequence 529, Application US/09770445
; Patent No. US2002023281A1

GENERAL INFORMATION:

; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Ted
; APPLICANT: Slader, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurdan, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 529
; LENGTH: 881
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-529

Alignment Scores:
Pred. No.: 3,96e-51 Length: 881
Score: 477.50 Matches: 96
Percent Similarity: 65.57% Conservative: 43
Best Local Similarity: 45.28% Mismatches: 55
Query Match: 43.21% Indels: 18
DB: 10 Gaps: 4

US-09-817-198A-2 (1-212) x US-09-770-445-529 (1-881)

QY 1 MetAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuGlyAspSerGlyValGly 20

Db 60 ATGAATCCTGAGTACGACTATCTTTCAAGCTCCTGCTTATCGGGATTCGTGGCTAGGC 119
QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisIleSerThr 40
Db 120 AAGTCCTGTCTCTTTGAGATCTCTGATGATCTTATGTAGAAGTGTACATTAGCACT 179
QY 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGln 60
Db 180 ATTGGAGTGCATTTAAATTTAGACTGTGGACACAGATGGCAACAATTAAGCTCCA 239
QY 61 IleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
Db 240 ATTTGGACACTGCTGCTCAAGAAGCTTCAGGACTATTACTAGCACTTACTACCGTGG 299
QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
Db 300 GCACATGGAAATATTATGTCTACGATGTCACAGATGAAGAAGCTTCAATATGTCAAG 359
QY 101 LysTrpValSerAspValAspGluTyrAlaProGluGluValGlnLysIleLeuIleGly 120
Db 360 CAATGTTGAGTGAATGTATGCTTATGCTAGTCAATGTCACAACTCCTTGTGGA 419
QY 121 AsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnLysLeuAla 140
Db 420 AACAAAGTCTCATCTTACTGAAACACAGAGCCATTCCTTATGAAACTGCCAAGGCTTTGCC 479
QY 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
Db 480 GATGAATCGGGATTCCTTTATGGAGACTAGTGCAGAAAGATGCTCAACAAGCTAGACAG 539
QY 161 SerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGlyLeu 180
Db 540 GCITTC-----ATGGCAATGCTGCATCCATCAAGAG-----572
QY 181 ArgMetArgAlaSer-AsnGluLeuAlaLeuAlaGluLeuGluGluGluGlyLysPr 200
Db 573 AGAATG---GCTAGCCACACAGCTGGGAATAATGCAA-----GA 608
QY 200 oGluGlyProAlaAsnSerSerLysThrCysTrp 211
Db 609 CCACCGACCGTGCAGATCATGAGGACGCCCTGTGG 642

RESULT 13
US-09-938-842A-832
; Sequence 832, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCDP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 832
; LENGTH: 609
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-832

Alignment Scores: 3.67e-50 Length: 609
Pred. No.: 468.00 Matches: 84
Score:

Percent Similarity: 71.01% Conservative: 36
Best Local Similarity: 49.70% Mismatches: 49
Query Match: 42.35% Indels: 0
DB: Gaps: 0
US-09-817-198A-2 (1-212) x US-09-938-842A-832 (1-609)
QY 1 MetalAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerGlyValGly 20
Db 1 ATGAATCCTGGAATGACTATTGTTCAAGCTTTGCTCATTTGGTGTGATTTCTGGTGTGGA 60
QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisIleSerThr 40
Db 61 AAGTCGTCTGCTTCTTAAGGTTTCTGATGATTCCTACCTGGATAGTACATCAGCACC 120
QY 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGln 60
Db 121 ATTGCTGTCGACTTTAAATCCGACACTCGAACAAGATGGAAGACCATCAAACTCCAG 180
QY 61 IleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
Db 181 ATTTGGATACGCGCAGGCGAGACGTTTTCAGACGATTACTAGCAGTTACTACAGAGA 240
QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
Db 241 GCTCATGGAAATTTATGTGACTTATGATGATGACGATCTAGAAAGCTTCAACAAGCTCAAG 300
QY 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
Db 301 CAATGGCTAAATGAATCGACCGCTATGCAAGTGAGATGTTTAAACAAGCTACTGCTGG 360
QY 121 AsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnGlnLeuAla 140
Db 361 AACAAAGTGTGCTCATCATCATCAGACAGAAAGTTGATCCACTGACAGACAGCTTTCGCT 420
QY 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
Db 421 GATGAATCTGGGATCCCATCTTGGAAACAAGTGTGAAGAATGCTTACCAATGCTCGAAGAA 480
QY 161 SerPheThrArgLeuThrGluLeuVal 169
Db 481 GCITTCATGCCCATGACTGCTGCAATC 507

RESULT 14
US-09-917-800A-1461
; Sequence 1461, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Portner, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884

```
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1461
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 U37099
US-09-917-800A-1461

Alignment Scores:
Pred. No.: 1,15e-47 Length: 585
Score: 448.00 Matches: 79
Percent Similarity: 69.89% Conservative: 44
Best Local Similarity: 44.89% Mismatches: 53
Query Match: 40.54% Indels: 0
DB: 10 Gaps: 0

US-09-817-198A-2 (1-212) x US-09-917-800A-1461 (1-585)

QY 3 LysGlnTyrAspValLeuPheArgLeuLeuLeuLeuGlyAspSerGlyValGlyLysThr 22
Db 4 CAGAATTTGATTGATGTTCAAGTTGCTGATCATTCATTCGCAATAGCAGCGTGGCAACA 63
QY 23 CysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisSerHisThrIleGly 42
Db 64 TCCTTCTGTTCCGCTATGCTGATGCTCTTCAGCTTCAGTCCCTTTGTCAGCAGCGTGGC 123
QY 43 ValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGlnIleTrp 62
Db 124 ATCGATTTCAAAGTAAATCTCTTCAAAATGAAAGAGAAATCAAGCTTCAGATTGG 183
QY 63 AspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgAlaGln 82
Db 184 GACAGCAGCGCAGGAAGATACAGGACCATCACACAGCCCTATTATCGAGGGGCCATG 243
QY 83 GlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMetLysTrp 102
Db 244 GCGTTCATTTTAAATGATGATACATCAATGAAGAAATCCCTCAACGCTGTACAGATTGG 303
QY 103 ValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGlyAsnLys 122
Db 304 TCAACTCAGATCAAAACATATTCCTGGGATATGCCAGGTATCTGCCCGGAACAAA 363
QY 123 AlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnGlnLysGlu 142
Db 364 TGTGACATGGAAGACGAGCGGTGCTCAACTGAGAGAGGCGCGCTTAGGAGAGCAG 423
QY 143 TyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGluSerPhe 162
Db 424 CTCGGGTTGAGTTTTTGAACACGCGCCAGGATACATCAAGCTCAAGCAACCTTT 483
QY 163 ThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGlu 178
Db 484 GAGCGCTCGTAGATATCATCTGTGACAAAATGTCAGAGCGTGGAG 531

RESULT 15
US-10-051-986-10
; Sequence 10, Application US/10051986
; Patent No. US20020146770A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; Tang, Y. Tom
; Lal, Preeti
; Guegler, Karl J.
; Corley, Neil C.
; Patterson, Chandra
; Batra, Sajeew
; Baughn, Mariah R.
; TITLE OF INVENTION: RAS PROTEINS
```

```
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
; CURRENT APPLICATION NUMBER: US/10/051,986
; FILING DATE: 15-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/766,551
; FILING DATE: DECEMBER 12, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0168-1 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0355
; TELEFAX: 650-845-4166
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 875 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: UCMCL5T01
; CLONE: 1528559
; SEQUENCE DESCRIPTION: SEQ ID NO: 10 :
US-10-051-986-10

Alignment Scores:
Pred. No.: 5.08e-45 Length: 875
Score: 429.00 Matches: 82
Percent Similarity: 68.48% Conservative: 44
Best Local Similarity: 44.57% Mismatches: 46
Query Match: 38.82% Indels: 12
DB: 12 Gaps: 3

US-09-817-198A-2 (1-212) x US-10-051-986-10 (1-875)

QY 5 TyrAspValLeuPheArgLeuLeuLeuGlyAspSerGlyValGlyLysThrCysLeu 24
Db 72 TAGGACCTCAGGGCAAGGTGATGCTCTGGGAGACACAGGCGTCGGCAACATGTTTC 131
QY 25 LeuCysArgPheThrAspAsnGluPheHisSer---SerHisIleSerThrIleGlyVal 43
Db 132 CTGATCCAATTCAAAGACGGGCGCTTCGTCGCGGAACCTTCATAGCCACCGTCGCGCAT 191
QY 44 AspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGlnIleTrpAsp 63
Db 192 GACTTCAGGAACAAGGTGCTGCTGCTGATGCGTGAAGTGAAGTGAAGTGAAGTGAAGT 251
QY 64 ThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgAlaGlnGly 83
Db 252 ACCGTGGCAGAACGCGTTCGGAAGCGTCACCCATGCTTATTACAGATGCTCAGGCC 311
QY 84 IlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMetLysTrpVal 103
Db 312 TTGCTTCTGCTGATGATGATCACCACAAATCTCTTTCACACACATCAGGCGCTGGCTC 371
QY 104 SerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGlyAsnLysAla 123
Db 372 ACTGAGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 431
```

```

QY 124 AspGluGluGlnLysArgGlnValGlyArgGluGlnGlyGlnGlnLeuAlaLysGluTyr 143
    ||| :||| :||| ||| |||:| |||:| |||:| |||:| |||:|
Db 432 GATATGAGCAGCGAAGAGTGTCCGTTCCGAGACGGAGACCTTGGCCAGGGAGTAC 491
    |||:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 144 GlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGluSerPheThr 163
    |||:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 492 GGTGTTCCCTTCCCTGGAGACCGCGCAAGACTGGCATGAATGTG----- 536
    |||:| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 164 ArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGlyLeuArgMetArg 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 -----GAGTTAGCCTTTCTGGCCATCGCCCAAGGAA-----CTGAAATACCGG 578
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184 AlaSerAsnGlu 187
    ||| :||| :|||
Db 579 GCCGGGCATCAG 590
    ||| :||| :|||

```

Search completed: November 17, 2002, 22:06:14
Job time : 55 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2002, 20:38:44 ; Search time 255 Seconds
(without alignments)
1872.250 Million cell updates/sec

Title: US-09-817-198a-2

Perfect score: 1105
Sequence: 1 MAKQYDVLFRLLIGSGVG.....LEEEGKPEGPANSSKTCWC 212

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+D2n.model -DEV=xlh
-O=/cgn2_1/USPRO_spool/US09817198/runat_13112002_135206_7022/app_query.fasta_1.391
-DB=N.Geneseq_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -DOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09817198 -CGCN_1_1_125_arunat_13112002_135206_7022 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1105	100.0	2021	22	AAS27053	CDNA encoding nove
2	1092	98.8	1666	21	AAC75813	Human ORFX ORF1368
3	832	75.3	566	22	AAS27472	CDNA encoding nove
4	546.5	49.5	1540	23	ABL27707	Drosophila melanog
5	545	49.3	1986	22	ABA09160	Human rab8 homolog
6	540	48.9	2411	24	ABQ54410	Human ovarian anti
7	534.5	48.4	1274	21	AAF18072	Lung cancer associ
8	530.5	48.0	716	21	AAA40104	Human Rab10 CDNA
9	530.5	48.0	861	21	AAA40108	Human Rab10 CDNA #
10	530.5	48.0	888	21	AAA40109	Canine Rab10 CDNA
11	530.5	48.0	956	21	AAAG6887	Nucleotide sequenc
12	530.5	48.0	1537	21	AAF16196	Human prostate can
13	530.5	48.0	3533	22	AAH17889	Human cDNA sequenc
14	529.5	47.9	674	23	AAS71453	DNA encoding novel
15	528.5	47.8	1161	22	AAH75182	Nucleotide sequenc
16	528.5	47.8	2247	23	ABL29661	Drosophila melanog
17	528.5	47.8	2497	22	AAS60878	Human cancer agent
18	528.5	47.8	2497	22	AAS60884	Human cancer agent
19	528.5	47.8	2497	22	AAS60893	Human cancer agent
20	528.5	47.8	2497	22	AAS60895	Human cancer agent
21	528.5	47.8	2497	23	ABV25781	Human prostate exp
22	528.5	47.8	2497	23	ABV30037	Human prostate exp
23	528.5	47.8	3077	22	AAH13912	Human cDNA sequenc
24	519.5	47.0	866	22	AAH04301	Human cDNA clone (
25	519	47.0	1023	21	AAC51491	Arabidopsis thalia
26	519	47.0	1025	21	AAC34347	Arabidopsis thalia
27	513.5	46.5	1129	21	AAC44482	Zea mays DNA fragm
28	510	46.2	911	21	AAC38429	Arabidopsis thalia
29	507	45.9	1101	21	AAC34080	Arabidopsis thalia
30	504	45.6	1203	21	AAC43685	Zea mays DNA fragm
31	501	45.3	705	21	AAC49208	Arabidopsis thalia
32	494	44.7	1153	21	AAC32781	Arabidopsis thalia
33	478.5	43.3	925	20	AAX37232	Human Rab protein,
34	478.5	43.3	939	21	AAF21661	Human breast and o
35	478.5	43.3	1898	22	AAS44924	CDNA encoding nove
36	478.5	43.3	1944	22	AAI93456	Human polynucleoti
37	477.5	43.2	730	21	AAZ36696	Nucleotide sequenc
38	477.5	43.2	777	21	AAC42684	Arabidopsis thalia
39	477.5	43.2	881	24	ABN98761	Arabidopsis thalia
40	472.5	42.8	607	22	AAH87926	Peppermint plant o
41	472.5	42.8	1202	21	AAF14053	Aspergillus oryzae
42	468	42.4	959	21	AAC33987	Arabidopsis thalia
43	467	42.3	723	24	ABK83978	Human cDNA differe
44	467	42.3	2528	21	AAC60009	Human secreted pro
45	466	42.2	932	21	AAC35200	Arabidopsis thalia

ALIGNMENTS

RESULT 1

ID AAS27053

XX AAS27053 standard; cDNA; 2021 BP.

AC AAS27053;

XX

DT 07-NOV-2001 (first entry)

XX CDNA encoding novel signal transduction pathway protein, Seq ID 88.

DE Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antinflammatory; anti-HIV; antibacterial; antinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;

PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-465460/50.
 DR P-PSDB; AAU17136.
 XX Novel polypeptides useful for diagnosing, treating, preventing and/or
 PT prognosing disorders related to the proteins, including cancers, immune
 PT disorders and neuronal disorders -
 XX Claim 1; SEQ ID No 88; 880pp; English.
 XX The invention relates to novel isolated polypeptides (I), and
 CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
 CC diagnosing, preventing and treating diseases including immune system
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
 CC transplant rejections and graft versus host disease, infectious diseases
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
 CC other blood-related disorders (sickle cell anaemia), myeloproliferative
 CC disorders, primary haematopoietic disorders, hyperproliferative
 CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
 CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders
 CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
 CC Addison's disease), reproductive system disorders, gastrointestinal
 CC disorder (inflammatory disorders), liver disorders (cirrhosis),
 CC as stimulators of B-cell responsiveness to pathogens, activators of
 CC T-cells, to induce higher affinity antibodies, and as a means to induce
 CC tumour proliferation in pathologies e.g. acquired immune deficiency
 CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
 CC pathway protein coding sequences and PCR primers of the invention.
 XX

Alignment Scores:
 Pred. No.: 6.03e-118 Length: 2021
 Score: 1105.00 Matches: 212
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0

US-09-817-198A-2 (1-212) x AAS27053 (1-2021)

QY 1 MetAlaLysGlnTyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerGlyValGly 20
 DB 81 ATGGCGAGACAGTACGATGCTGTTCGGCTGCTGTCGTGATCGGGAGCTCCGGGTGGGC 140
 QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisIleSerThr 40
 DB 141 AAGACCTGCTGCTGCGGCTTACCGACAACGAGTTCCTCTCGCACATCTCCACC 200
 QY 41 IleGlyValAspPheIleGlyMetIleThrIleGluValAspGlyIleIleValArgIleGln 60
 DB 201 ATCGGTGTGATTTAAGATGAAGACCATAGAGGTAGACGGCATCAAGTCCGATACAG 260
 QY 61 IleTyrAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
 DB 261 ATCTGGACATCTGACGGCAGGAGATACACAGACCATCAACAGAGTACTATCGGGG 320
 QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
 DB 321 GCCACGGGATATTTTGGTCTATGACATATACGACGAGCGCTCTTACCAGCACATCATG 380
 QY 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
 DB 381 AAGTGGGTGAGTGTGATGAGTACGCACCAAGGCGTCCAGAGATCTTATTGGG 440

QY 121 AsnLysAlaAspGluGluGlnLysArgGlnValGlyArgGluGlnGlnGlnLeuAla 140
 DB 441 AATAAGGCTGATGAGGACAGAAACGGCAGGTGGGAAGAGACCAAGGGCAGCTGGCG 500
 QY 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
 DB 501 AAGGAGTATGGCATGCACTTCTATGAACACAGTGGCTGCACCAACCTCAACATTAAAGAG 560
 QY 161 SerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGlyLeu 180
 DB 561 TCATTACGCGTCTGACAGAGCTGGTGTGAGGCCCATAGGAAGAGCTGGAAGGCCCTC 620
 QY 181 ArgMetArgAlaSerAsnGluLeuAlaLeuAlaGluLeuGluGluGluGlyLysPro 200
 DB 621 CGGATCGCTGCCAGCAATGAGTTGGCACTGGCAGAGCTGGAGGAGGAGGAGGAGGAG 680
 QY 201 GluGlyProAlaAsnSerSerLysThrCysTrpCys 212
 DB 681 GAGGCGCCAGCGCACTTTCGAAACCTGCTGGTGC 716
 RESULT 2
 AAC75813
 ID AAC75813 standard; cDNA; 1666 BP.
 AC AAC75813;
 XX 08-FEB-2001 (first entry)
 DE Human ORFX ORF1368 polynucleotide sequence SEQ ID NO:2735.
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipariatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX Homo sapiens.
 OS
 XX WO200058473-A2.
 PN
 XX 05-OCT-2000.
 PD
 XX 31-MAR-2000; 2000WO-US08621.
 PF
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX (CURA-) CURAGEN CORP.
 PA
 XX Shimkets RA, Leach M;
 PI
 XX WPI; 2000-602362/57.
 DR P-PSDB; AAB41604.
 XX Novel nucleic acids and peptides derived from open reading frame x,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX Claim 5; Page 1978-1979; 5507pp; English.
 PS AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

QY 21 LysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisIleSerThr 40
DB 131 AAGACCTGCTGCTGTGCGCTTCCCGACAAACGAGTTTCCACTCTCGCACATCTCCACC 190
QY 41 IleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIleGln 60
DB 191 ATCGGTGTGACTTTAAGATGAAGACCATAGAGGTAGACGGCATCAAAAGTGGGATACAG 250
QY 61 IleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArgArg 80
DB 251 ATCTGGACACTGCAGCGCAGGAGAGATACACAGCATCACAAGCAGTACTATCGCGG 310
QY 81 AlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIleMet 100
DB 311 GCCCAGGGGATATTTTGGTCTATGATCATAGCAGGCGGCTCTTACCAGCACATCATG 370
QY 101 LysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIleGly 120
DB 371 AAGTGGGTCACTGACGTGGATGAGTACGCCACCCAGAGGCGTCCAGAAATCCTTATTGGG 430
QY 121 AsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnGlnLeuAla 140
DB 431 AATAAGGCTGATGAGGAGCAGAWACGCCAGGTGGGAAGAGAGCAAGGGCAGCANCTGGCG 490
QY 141 LysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLysGlu 160
DB 491 AANGAGTATGGCATGGACTTCTATGAACAAAGTGCCTGCACCAACCTNACATTAAAGAG 550
QY 161 SerPheThrArg 164
DB 551 TCATTCACGCGT 562
RESULT 4
ABL27707
ID ABL27707 standard; DNA; 1540 BP.
AC ABL27707;
XX
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 34594.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW Drosophila; gene; ds.
KW Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 34594; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1540 BP; 437 A; 353 C; 420 G; 330 T; 0 other;
Alignment Scores: Length: 1540
Pred. NO.: 2.19e-53 Matches: 105
Score: 546.50 Conservativity: 47
Percent Similarity: 72.04% Mismatches: 50
Best Local Similarity: 49.76% Indels: 9
Query Match: 49.46% Gaps: 3
DB: 23
US-09-817-198A-2 (1-212) x ABL27707 (1-1540)
QY 1 MetAlaLysGln---TyrAspValLeuPheArgLeuLeuLeuLeuLeuLeuLeuVal 19
DB 257 ATGGCAAGAAACCTACGATTGCTTTAAACTGTTGCTGATCGGTGATTTCAGAGTG 316
QY 20 GlyLysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisIleSer 39
DB 317 GCGAAGACGTGATATTTGTCGGGTTCGGATGATGATTCACGTCCACGTTCATATCG 376
QY 40 ThrIleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIle 59
DB 377 ACCATAGGCATCGATTTCAAAATCAAAACAGTCGAGTCGCGGCAAGAGATCAAGCTG 436
QY 60 GlnIleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArg 79
DB 437 CAAATATGGACACCGCGCGCAGGAGCGGTTCACACAGCATAAACACCTCGTACTATCGA 496
QY 80 ArgAlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIle 99
DB 497 GCGCCATGGCATATGCTGTCTATGACATAACAGAGAGAGATTCGAGAACATA 556
QY 100 MetLysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIle 119
DB 557 GTCAATGTTACGGAATATAGACAGACGCGCAACAGGAGATGTGGAGAAGATGATCCTC 616
QY 120 GlyAsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnGlnLeu 139
DB 617 GGCACAAGTGCATATGACGGACAAGAGGTGGTCAACAAGGAGCGCGCGAAGCGATT 676
QY 140 AlaLysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLys 159
DB 677 GCCCGTGACATGGCATTCGGTTTATGGAACATCCGCCAAGTCGAACATAACATCGAG 736
QY 160 GluSerPheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLysGluLeuGluGly 179
DB 737 CGGCGCTTCTCGAGCTGGCGCGAGCCATCTCGAC-----AAGACATCAGGA 784
QY 180 LeuArgMetArgAlaSerAsnGluLeuAlaLeuAlaGluLeuGluGluGlyLys 199
DB 785 CGCGAGTCGCGGAGATCAGGAGCGGTGATTATTCATCGCCGGAACCGAGGAGCGG 844
QY 200 ProGluGlyProAlaAsnSerSerLysThrCys 210
DB 845 CCG-----GGCTACAGCAAGTGTGCTGC 865
RESULT 5
ABA09160
ID ABA09160 standard; cDNA; 1986 BP.
XX
XX ABA09160;
AC
XX
XX 11-JAN-2002 (first entry)
DT
XX Human rab8 homologue-encoding cDNA, SEQ ID NO:936.
DE


```
Db 198 ACCATAGGATGACATTCAGATCAAAACAGTTGATTACAAAGGAAGATCAAGCTA 257
Qy 60 GlnileTirPaspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArg 79
Db 258 CAGATATGGATACAGCAGCCAGGAGCGATTTCACACCATCACACACCTCTACTACAGA 317
Qy 80 ArgAlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIle 99
Db 318 GCGCGCAATGGGTATCATGCTAGTATATGACATCACCACATGGTAAAGTTTGGAAACATC 377
Qy 100 MetLysTirPValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIle 119
Db 378 AGCAATGGCTTAGAAACATAGATGAGCATGCCAATGAAGATGTGAAAGAAATGTTACTA 437
Qy 120 GlyAsnLysAlaAspGluGlnGlnLysArgGlnValGlyArgGluGlnGlnGlnLeu 139
Db 438 GGAACAAAGTGTATATGGACGACAAAGAGTGTACCTAAAGGAAAGAGAGAACAGATT 497
Qy 140 AlaLysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLys 159
Db 498 GCRAAGGAGCATGGTATTAGCTTTTGTAGACTAGTGCRAAAGCAAATATAACATCGAA 557
Qy 160 GluSerPheThrArgLeuThrGluLeuValLeuGln 171
Db 558 AAGCGCTTCTCAGCTTAGCTGAAGATATCCTTCGA 593
RESULT 9
AA440108
ID AAA40108 standard; cDNA; 861 BP.
XX AC AAA40108;
XX XX
XX XX
DT 19-OCT-2000 (first entry)
DE Human Rab10 cDNA #2.
XX Human; Rab10; Yap/Rab family; ss.
XX Homo sapiens.
XX CN1249345-A.
XX PD 05-APR-2000.
XX PF 28-SEP-1998; 98CN-0121911.
XX PR 28-SEP-1998; 98CN-0121911.
XX PA (UYFU-) UNIV FUDAN.
XX Yu L, Tu Q, Gao J;
XX WPI; 2000-400723/35.
XX Preparation of human gene coding sequence, its encoded polypeptide -
XX Disclosure; Fig 1; 23pp; Chinese.
CC This invention describes a novel human Rab10 cDNA sequence. The protein
CC coded by said sequence belongs to Yap/Rab family and is the homolog of
CC mouse Rab10. The present invention also relates to the polypeptide coded
CC by said nucleotide sequence and the application and preparing process
CC of said polynucleotide and said polypeptide. This sequence encodes the
CC human Rab10 protein described in the method of the invention.
XX
SQ Sequence 861 BP; 258 A; 189 C; 200 G; 214 T; 0 other;
Alignment Scores:
Pred. No.: 7.08e-52 Length: 861
Score: 530.50 Matches: 96
Percent Similarity: 82.56% Conservative: 46
Best Local Similarity: 55.81% Mismatches: 29
Query Match: 48.01% Indels: 1
1 2
```

```
DB: 21 Gaps: 1
US-09-817-198a-2 (1-212) x AAA40108 (1-861)
Qy 1 MetAlaLysGln---TyrAspValLeuPheArgLeuLeuLeuIleGlyAspSerGlyVal 19
Db 95 ATGGCGCAAGAACAGACCTACGACCTGCTTTCAAGCTGCTCCCTGATCGGGATTTCCGGAGTG 154
Qy 20 GlyLysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisIleSer 39
Db 155 GGAAGACCTGGCTCCTTTTCGTTTCGGATGATGCTTCAATACTACCTTTATTTCC 214
Qy 40 ThrIleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIle 59
Db 215 ACCATAGGAATAGACTTCAAGATCAAAACAGTTGAATTACAAAGGAAGATCAAGCTA 274
Qy 60 GlnileTirPaspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArg 79
Db 275 CAGATATGGATACAGCAGCCAGGAGCGATTTCACACCATCACACCTCTACTACAGA 334
Qy 80 ArgAlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIle 99
Db 335 GCGCGCAATGGGTATCATGCTAGTATATGACATCACCACATGGTAAAGTTTGGAAACATC 394
Qy 100 MetLysTirPValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIle 119
Db 395 AGCAATGGCTTAGAAACATAGATGAGCATGCCAATGAAGATGTGAAAGAAATGTTACTA 454
Qy 120 GlyAsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnGlnLeu 139
Db 455 GGAACAAAGTGTATATGGACGACAAAGAGCTGTACTTAAAGGAAGAGAGAACAGATT 514
Qy 140 AlaLysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLys 159
Db 515 GCAAGGAGCATGGTATTAGCTTTTGTAGACTAGTGCRAAAGCAAATATAACATCGAA 574
Qy 160 GluSerPheThrArgLeuThrGluLeuValLeuGln 171
Db 575 AAGCGCTTCTCAGCTTAGCTGAAGATATCCTTCGA 610
RESULT 10
AA440109
ID AAA40109 standard; cDNA; 888 BP.
XX AC AAA40109;
XX XX
DT 19-OCT-2000 (first entry)
DE Canine Rab10 cDNA.
XX Human; Rab10; Yap/Rab family; canine; ss.
XX Canis sp.
XX CN1249345-A.
XX PD 05-APR-2000.
XX PF 28-SEP-1998; 98CN-0121911.
XX PR 28-SEP-1998; 98CN-0121911.
XX PA (UYFU-) UNIV FUDAN.
XX Yu L, Tu Q, Gao J;
XX WPI; 2000-400723/35.
XX Preparation of human gene coding sequence, its encoded polypeptide -
XX Disclosure; Fig 1; 23pp; Chinese.
XX This invention describes a novel human Rab10 cDNA sequence. The protein
```


CC coded by said sequence belongs to Yap/Rab family and is the homolog of
CC mouse Rab10. The present invention also relates to the polypeptide coded
CC by said nucleotide sequence and the application and preparing process
CC of said polynucleotide and said polypeptide. This sequence encodes a
CC canine Rab10 protein described in the method of the invention.

XX
SQ Sequence 888 BP; 261 A; 202 C; 202 G; 223 T; 0 other;

Alignment Scores:
Pred. No.: 7.38e-52 Length: 888
Score: 530.50 Matches: 96
Percent Similarity: 82.56% Conservative: 46
Best Local Similarity: 55.81% Mismatches: 29
Query Match: 48.01% Indels: 1
DB: 21 Gaps: 1

US-09-817-198A-2 (1-212) x AAA40109 (1-888)

QY 1 MetAlaLysGln---TyrAspValLeuPheArgLeuLeuLeuLeuLeuLeuLeuVal 19
Db 133 ATGGCGAAGAGACGACGCTGCTTTTCAAGCTGCTCTGATCGGAGACTCGGGAGTA 192
QY 20 GlyLysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisLeuSer 39
Db 193 GGGAAAGACCTGGCTCTTTTTCGGATGATGCTTCAATACCACCTTTATTTC 252
QY 40 ThrIleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIle 59
Db 253 ACCATAGGAATAGATTAAAGATCAAAACAGTTGAATTACAGGAAAGATCAAGCTA 312
QY 60 GlnIleTyrAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrArg 79
Db 313 CAGATATGGATACAGCAGGCGAGGATTCACACCATCAACACCTCTACTACAGA 372
QY 80 ArgAlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIle 99
Db 373 GGAGCAATGGGTATCATGCTAGTATATGACATCAACCAATGGTAAAGTTTGAACATC 432
QY 100 MetLysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIle 119
Db 433 ACCAATGGCTTAGAAACATAGATGAGTGCCTAAATGAAGATGTGGAAGATGTACTA 492
QY 120 GlyAsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnGlnGln 139
Db 493 GCAACAAATGATGATGACGACATAAAGAGTTGTACCTAAAGGAAGAGGACAGATT 552
QY 140 AlaLysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLys 159
Db 553 GCAAGGGAGCATGGTATTAGATTTTTCGAGACTAGTGCAGAAAGATTAATATAACATCGAA 612
QY 160 GluSerPheThrArgLeuThrGluLeuValLeuGln 171
Db 613 AAGGCTTCCTCACATTACCTGAAGATATCTTCGA 648

RESULT 11

AAA96887

ID AAA96887 standard; DNA; 956 BP.

XX

AC AAA96887;

XX

DT 19-FEB-2001 (first entry)

XX

DE Nucleotide sequence of human RAB10.

XX

KW RAB protein; GTPase; GTP binding; gene therapy; cancer; ss.

XX

OS Homo sapiens.

XX

EH Key Location/Qualifiers

FT CDS 127..729

FT /*tag= a

FT /*product= "RAB10"

XX

PN WO200058464-A2.

XX

PD 05-OCT-2000.

XX

PF 13-MAR-2000; 2000WO-US06330.

XX

PR 25-MAR-1999; 99US-0126083.

XX

PA (AXYS-) AXYS PHARM INC.

XX

PI Allen M, Abel K, McIntosh B, Vega R, Rutter M, Buckler AJ;

XX

DR WPI; 2000-647233/62.

XX

DR P-PSDB; AAB19165.

XX

PT Novel isolated nucleic acid encoding a mammalian RAB protein useful for
PT identifying homologous or related genes, in producing composition that
PT modulates expression or function of RAB for cancer therapy -

XX

PS Claim 4; Page 38-39; 58pp; English.

XX

CC The present sequence encodes a mammalian RAB protein. RAB proteins
CC constitute the largest family of small GTPases, with over 40 identified
CC isoforms. RAB proteins contain four highly conserved peptide sequences
CC involved in GTP binding and hydrolysis. Compositions comprising RAB
CC nucleic acid are useful for identifying homologous or related genes,
CC in producing compositions that modulate the expression or function of
CC RAB, for gene therapy, mapping functional regions of the protein and
CC in studying associated physiological pathways. In addition, modulation
CC of the gene activity in vivo is used for prophylactic and therapeutic
CC purposes, such as treatment of cancer, and identification of cell type
CC based on expression. The DNA may also be used to identify expression of
CC the gene in a biological specimen.

XX SQ Sequence 956 BP; 294 A; 205 C; 213 G; 244 T; 0 other;

Alignment Scores:

Pred. No.: 8.16e-52 Length: 956
Score: 530.50 Matches: 96
Percent Similarity: 82.56% Conservative: 46
Best Local Similarity: 55.81% Mismatches: 29
Query Match: 48.01% Indels: 1
DB: 21 Gaps: 1

US-09-817-198A-2 (1-212) x AAA96887 (1-956)

QY 1 MetAlaLysGln---TyrAspValLeuPheArgLeuLeuLeuLeuLeuLeuVal 19

Db 127 ATGGCGAAGAGACGACGCTGCTTTTCAAGCTGCTCTGATCGGAGACTCGGGAGTG 186

QY 20 GlyLysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerHisLeuSer 39

Db 187 GGGAAAGACCTGGCTCTTTTTCGGATGATGCTTCAATACCTATTATTTC 246

QY 40 ThrIleGlyValAspPheLysMetLysThrIleGluValAspGlyIleLysValArgIle 59

Db 247 ACCATAGGAATAGACTTCAAGATCAAAACAGTTGAATTACAGGAAGAGATCAAGCTA 306

QY 60 GlnIleTyrAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrArg 79

Db 307 CAGATATGGATACAGCAGGCGAGGATTCACACCATCAACACCTCTACTACAGA 366

QY 80 ArgAlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIle 99

Db 367 GCGCAATGGGTATCATGCTAGTATGACATCAACCAATGGTAAAGTTTGAACATC 426

QY 100 MetLysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIle 119

Db 427 ACCAATGGCTTAGAAACATAGATGCGCATGAGATGTGGAAGATGTACTA 486

QY 120 GlyAsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnGlnGln 139

Db 487 GGAACAAAGTGTGATGGACGACAAAGAGTTGTACCTAAAGGAAGAGGACAGAT 546

PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
PS Claim 8; SEQ ID 17618; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX

SQ Sequence 3533 BP; 967 A; 716 C; 730 G; 1120 T; 0 other;

Alignment Scores:

Pred. No.:	4.85e-51	Length:	3533
Score:	530.50	Matches:	96
Percent Similarity:	82.56%	Conservative:	46
Best Local Similarity:	55.81%	Mismatches:	29
Query Match:	48.01%	Indels:	1
DB:	22	Gaps:	1

US-09-817-198A-2 (1-212) x AAH17889 (1-3533)

Qy	1	MetAlaLysGln---TyrAspValLeuPheArgLeuLeuLeuLeuGlyAspSerGlyVal	19
Db	500	ATGGCGAAGAGACGTACCGACTGCTTTCAAGCTGCTCTGATCGGGGATTCGGAGTG	559
Qy	20	GlyLysThrCysLeuLeuCysArgPheThrAspAsnGluPheHisSerSerHisLeuSer	39
Db	560	GGGAAGACCTCGCTCTTTTTCGTTTTCGGATGATGCTTCAATACCTATTTATTTCC	619
Qy	40	ThrIleGlyValAspPheLysMetLysThrIleGluValAspGlyLeuLysValArgIle	59
Db	620	ACCATAGGAAATGACTTCAAGATCAAAACAGTTGAAATACAGGAAGAAGATCAAGCTA	679
Qy	60	GlnIleTrpAspThrAlaGlyGlnGluArgTyrGlnThrIleThrLysGlnTyrTyrArg	79
Db	680	CAGATATGGATACAGCAGGCGAGGCGATTTACACACCATCACACCTCTACTACAGA	739
Qy	80	ArgAlaGlnGlyIlePheLeuValTyrAspIleSerSerGluArgSerTyrGlnHisIle	99
Db	740	GGCGAATGGGTATCATCTAGTATATGACATCACCAATGGTAAAGTTTGAAACATC	799
Qy	100	MetLysTrpValSerAspValAspGluTyrAlaProGluGlyValGlnLysIleLeuIle	119
Db	800	ACCAATGGCTTAGAACATAGATGAGCATGCCAATGAAGATGGGAAGAATGTTACTA	859
Qy	120	GlyAsnLysAlaAspGluGlnLysArgGlnValGlyArgGluGlnGlnGlnLeu	139
Db	860	GGAACAAAGTGTGATATGACACACAAAAGAGTTGTACCTAAAGGAAGAGGACAGAT	919
Qy	140	AlaLysGluTyrGlyMetAspPheTyrGluThrSerAlaCysThrAsnLeuAsnIleLys	159
Db	920	GCAAGGGAGCATGGTATTTAGTTTTCGAGACTAGTGCAAAAAGCAAAATATAACATCGAA	979

Qy	160	GlusPheThrArgLeuThrGluLeuValLeuGln	171
Db	980	AAGGGTTCCTCACGTTAGCTAGCAAGATATCCTTCGA	1015

RESULT 14

AA571453	
ID	AA571453 standard; cDNA; 674 BP.
XX	
AC	AA571453;
XX	
DT	13-FEB-2002 (first entry)
XX	
DE	DNA encoding novel human diagnostic protein #7257.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX	
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
XX	30-MAR-2001; 2001WO-US08631.
PF	
XX	
PR	31-MAR-2000; 2000US-0540217.
XX	
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmanac RT, Liu C, Tang YT;
XX	
DR	
DR	WPI: 2001-639362/73.
DR	P-FSDB; ABG07266.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	

Claim 1; SEQ ID No 7257; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AA564197-AA594564 represent novel human diagnostic coding sequences of the invention.
Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 674 BP; 218 A; 136 C; 173 G; 147 T; 0 other;

Alignment Scores:

Pred. No.:	6.62e-52	Length:	674
Score:	529.50	Matches:	96
Percent Similarity:	82.56%	Conservative:	46
Best Local Similarity:	55.81%	Mismatches:	29
Query Match:	47.92%	Indels:	1

Db 438 ATTGACTATGGGATTAAATCTTTGGAGACAAGCGCAAAATCCAGTGCAAATGTAGAAGAG 497

Qy 161 Ser---PheThrArgLeuThrGluLeuValLeuGlnAlaHisArgLys 175

Db 498 GCATTTTACACTTGCACGAGATATAATGACAAAACCTCAACAGAAAA 545

Search completed: November 17, 2002, 20:47:19
Job time : 260 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2002, 20:38:29 ; Search time 20 seconds

(without alignments)
159.645 Million cell updates/sec

Title: US-09-817-198A-2

Perfect score: 1105

Sequence: 1 MAKQYDVLFRLLIGDSGVG.....LEEEGKPEGPANSSKTCWC 212

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1105	100.0	212	US-09-817-198A-2	Sequence 2, Appli
2	1092	98.8	218	US-09-817-198A-5	Sequence 5, Appli
3	1077	97.5	212	US-09-817-198A-4	Sequence 4, Appli
4	532	48.1	246	US-09-925-302-534	Sequence 534, App
5	530.5	48.0	218	US-09-925-300-1571	Sequence 1571, Ap
6	528.5	47.8	207	US-09-794-257-8	Sequence 8, Appli
7	512	46.3	190	US-09-822-860-5	Sequence 5, Appli
8	497	45.0	162	US-09-834-765-766	Sequence 766, App
9	478.5	43.3	201	US-09-967-736-3	Sequence 3, Appli
10	469	42.4	201	US-09-967-736-8	Sequence 8, Appli
11	467	42.3	222	US-09-820-003A-4	Sequence 4, Appli
12	434.5	39.3	198	US-09-794-257-16	Sequence 16, Appl
13	434.5	39.3	198	US-09-945-173-5	Sequence 5, Appli
14	434.5	39.3	198	US-09-972-529-4	Sequence 4, Appli
15	430	38.9	223	US-09-817-199A-4	Sequence 4, Appli
16	429	38.8	223	US-09-817-199A-2	Sequence 2, Appli
17	416	37.6	191	US-09-794-257-14	Sequence 14, Appl
18	416	37.6	191	US-10-051-986-3	Sequence 3, Appli
19	412.5	37.3	212	US-09-350-874-67	Sequence 67, Appl

20	399	36.1	312	10	US-09-925-302-783	Sequence 783, App
21	376	34.0	213	10	US-09-794-257-5	Sequence 5, Appli
22	373	33.8	213	10	US-09-988-974-8	Sequence 8, Appli
23	373	33.8	217	10	US-09-988-974-3	Sequence 3, Appli
24	373	33.8	239	10	US-09-925-301-1077	Sequence 1077, Ap
25	370	33.5	201	10	US-09-822-860-2	Sequence 2, Appli
26	353.5	32.0	624	10	US-09-834-765-5	Sequence 5, Appli
27	353.5	32.0	625	10	US-09-834-765-762	Sequence 762, App
28	353.5	32.0	832	10	US-09-834-765-2	Sequence 2, Appli
29	351	31.8	168	10	US-09-834-765-765	Sequence 765, App
30	343	31.0	161	10	US-09-834-765-763	Sequence 763, App
31	339	30.7	208	9	US-10-108-605-45	Sequence 45, Appl
32	338.5	30.6	216	10	US-09-945-173-10	Sequence 10, Appl
33	338.5	30.6	217	10	US-09-925-300-1364	Sequence 1364, Ap
34	323.5	29.3	211	12	US-10-051-986-6	Sequence 6, Appli
35	321	29.0	173	10	US-09-820-003A-2	Sequence 2, Appli
36	320	29.0	259	12	US-10-051-986-1	Sequence 1, Appli
37	320	29.0	260	12	US-10-051-986-4	Sequence 4, Appli
38	310	28.1	157	10	US-09-834-765-764	Sequence 764, App
39	303.5	27.5	154	10	US-09-822-860-4	Sequence 4, Appli
40	300.5	27.2	106	10	US-09-867-550-1812	Sequence 1812, Ap
41	297	26.9	201	10	US-09-988-974-5	Sequence 5, Appli
42	297	26.9	201	10	US-09-988-974-9	Sequence 9, Appli
43	297	26.9	209	10	US-09-864-761-42996	Sequence 42996, A
44	290	26.2	144	10	US-09-972-529-7	Sequence 7, Appli
45	282.5	25.6	208	10	US-09-925-302-629	Sequence 629, App

ALIGNMENTS

RESULT 1

US-09-817-198A-2

; Sequence 2, Application US/09817198A

; Patent No. US20020146758A1

; GENERAL INFORMATION:

; APPLICANT: YE, Jane et al.

; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE

; TITLE OF INVENTION: PROTEINS, AND USES THEREOF

; FILE REFERENCE: CLO01188

; CURRENT APPLICATION NUMBER: US/09/817,198A

; CURRENT FILING DATE: 2001-03-27

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 212

; TYPE: PRT

; ORGANISM: Human

US-09-817-198A-2

Query Match 100.0%; Score 1105; DB 10; Length 212;
Best Local Similarity 100.0%; Pred. No. 1.2e-102;
Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAKOYDVLFRLLIGDSGVGKTCCLRCFTDNEFHSHSHSTIGVDFKMKTIYVDGKVRQ	60
Db	1	MAKOYDVLFRLLIGDSGVGKTCCLRCFTDNEFHSHSHSTIGVDFKMKTIYVDGKVRQ	60
Qy	61	IWDTAGQERYQTITTKOYRRAGQIFLVYDISSERSYQHIMKWSVDYDEYAPGVOKILIG	120
Db	61	IWDTAGQERYQTITTKOYRRAGQIFLVYDISSERSYQHIMKWSVDYDEYAPGVOKILIG	120
Qy	121	NKADEQKQVREGQOQALAKYGMDFYTSACTNLNIKESFRTLTELVLQHRKELEGL	180
Db	121	NKADEQKQVREGQOQALAKYGMDFYTSACTNLNIKESFRTLTELVLQHRKELEGL	180
Qy	181	RMRSNELALAELEEEGKPEGPANSSKTCWC	212
Db	181	RMRSNELALAELEEEGKPEGPANSSKTCWC	212

RESULT 2

US-09-817-198A-5
; Sequence 5, Application US/09817198A
; Patent No. US20020146758A1
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001188
; CURRENT APPLICATION NUMBER: US/09/817,198A
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-817-198A-5

Query Match 98.8%; Score 1092; DB 10; Length 218;
Best Local Similarity 97.2%; Pred. No. 2.4e-101;
Matches 212; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

QY 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSSHISTIGVDFKMKTIIEVDGKVRQ 60
DB 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSSHISTIGVDFKMKTIIEVDGKVRQ 60

QY 61 IWDTAGQERYQTITKQYRRAQGIFLVYDISSERSYQHIMKWSVDVDEYAPGVQKILIG 120
DB 61 IWDTAGQERYQTITKQYRRAQGIFLVYDISSERSYQHIMKWSVDVDEYAPGVQKILIG 120

QY 121 NKADBEQKRVGREOGQQLAKEYGMDFYETSACTNLIKESFTRLTVELVLAHRKEGL 174
DB 121 NKADBEQKRVGREOGQQLAKEYGMDFYETSACTNLIKESFTRLTVELVLAHRKEGL 180

QY 175 KELEGLRMRASNELALAELEEEGKPEGPANSSKTCWC 212
DB 181 KELEGLRMRASNELALAELEEEGKPEGPANSSKTCWC 218

RESULT 3
US-09-817-198A-4
; Sequence 4, Application US/09817198A
; Patent No. US20020146758A1
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001188
; CURRENT APPLICATION NUMBER: US/09/817,198A
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-817-198A-4

Query Match 97.5%; Score 1077; DB 10; Length 212;
Best Local Similarity 97.6%; Pred. No. 7.1e-100;
Matches 207; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSSHISTIGVDFKMKTIIEVDGKVRQ 60
DB 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSSHISTIGVDFKMKTIIEVDGKVRQ 60

QY 61 IWDTAGQERYQTITKQYRRAQGIFLVYDISSERSYQHIMKWSVDVDEYAPGVQKILIG 120
DB 61 IWDTAGQERYQTITKQYRRAQGIFLVYDISSERSYQHIMKWSVDVDEYAPGVQKILIG 120

QY 121 NKADBEQKRVGREOGQQLAKEYGMDFYETSACTNLIKESFTRLTVELVLAHRKEGL 180

DB 121 NKADBEQKRVGREOGQQLAKEYGMDFYETSACTNLIKESFTRLTVELVLAHRKEGL 180

QY 181 RMRASNELALAELEEEGKPEGPANSSKTCWC 212

DB 181 RTCASNELALAELEDEGKTEGPAANSSKTCWC 212

RESULT 4
US-09-925-302-534
; Sequence 534, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 534
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-302-534

Query Match 48.1%; Score 532; DB 10; Length 246;
Best Local Similarity 47.6%; Pred. No. 1.3e-45;
Matches 100; Conservative 49; Mismatches 43; Indels 18; Gaps 3;

QY 1 MAKQYDVLFRLLIGDSGVGKTCCLCRFTDNEFHSSHISTIGVDFKMKTIIEVDGKVRQ 60

DB 44 MAKAYDHLFRLLIGDSGVGKTCCLIRFADNENNYISTIGIDFKIRTVDSGKKIKLO 103

QY 61 IWDTAGQERYQTITKQYRRAQGIFLVYDISSERSYQHIMKWSVDVDEYAPGVQKILIG 120

DB 104 VWDTAGQEREFTITATYIRGAMGILVYDITDEKSFENIQNMKSIKENASAGVERILLG 163

QY 121 NKADBEQKRVGREOGQQLAKEYGMDFYETSACTNLIKESFTRLTVELVLAHRKEGL 180

DB 164 NKCDMEAKRVQEQADKLAREHGIRFFETSASKSMNVDEAFSSLDIL-----LKSG 217

QY 181 RMRASNELALAELEEEGKPEGPANSSKTC 210

DB 218 GRRSGN-----GNKP--PSTDLTCTC 235

RESULT 5
US-09-925-300-1571
; Sequence 1571, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1571
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1571


```

Query Match      43.38; Score 478.5; DB 10; Length 201;
Best Local Similarity 45.08; Pred. No. 2e-40;
Matches 91; Conservative 64; Mismatches 64; Indels 5; Gaps 1;

QY   1 MAKQDYVLFRLLLIGDSGVGKTCCLCRFTDNEFHSSHISTIGVDFFMKTIYEDVGIKVRIQ 60
      | : || : ||| ||||| ||| ||| : : | : ||||| ||| : ||| : || : || : ||
Db    1 MPEDYDFLKKLLIGDSGVGSCLLLRFADDTYESIVTIGVDFKIRTIELDGKTIKLQ 60

QY   61 IWDTAGQERYOTTKOYYRRAGGLFVYDISSERSYOHIMKWVSVDVEYAPEGVOKILIG 120
      | : ||| ||| ||| ||| ||| ||| : : | : ||||| ||| : ||| : || : || : ||
Db    61 IWDTAQGERFTITSSYIRGAHGIIYVDVTDQESYANVRQWLQEIDRYASENVNKLVLG 120

QY   121 NKADEEQKVQREGOGQLAKEXYGNDFVETSACINLNKIESEFRFLTELVLQAHRKELEG 180
      | : ||| : ||| : ||| ||| ||| ||| : : ||| : ||| : ||| : ||| : |||
Db    121 NKSLDTTKKVVNDNTAKEFAUSLGHPFLETSKANATNQEARF-----MTMAAEIKRMGP 175

QY   181 RMRASNELALALEEEEGKPEG 202
      : | : |||
Db    176 GAASGGERNLKIDSTPVKPAG 197

RESULT 10
US-09-967-736-8
; Sequence 8, Application US/09967736
; Patent No. US20020103340A1
; GENERAL INFORMATION:
; APPLICANT: Hallman, Jennifer L.
;              Lal, Preeti
;              Corley, Neil C.
;              Shah, Purvi
; TITLE OF INVENTION: RAB PROTEINS

```

[illegible]

```

RESULT 11
US-09-820-003A-4
; Sequence 4, Application US/09820003A
; Patent No. US20020142382A1
; GENERAL INFORMATION:
; APPLICANT: MERRULOV, Gennady et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001196
; CURRENT APPLICATION NUMBER: US/09/820,003A
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Homo sapien

```

US-09-820-003A-4

Query Match 42.3%; Score 467; DB 10; Length 222;
Best Local Similarity 51.9%; Pred. No. 3.2e-39;
Matches 84; Conservative 34; Mismatches 44; Indels 0; Gaps 0;

Qy 1 MAKQYDLVFRLLIGDSVGKTCLLCRFTDNEFHSHSTIGVDFKMTIEVDGKIVRIQ 60
I :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db 21 MNPEYDYLFKLLIGDSVGKSCLLRFDADDTYTESYISTIGVDFKIRTIELDGKTIKQ 80

Qy 61 IWDTAGQERYOTIKQYRRAGGIFLVYDITSSERSYQHIMKWSDVDVEYAPGVQKILIG 120
I :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db 81 IWDTAGQERFRITSSYRGAHGIIVYDVTQDSFNFVKWQLQEIIRHADKDNVPIVLVGNKCDLED 140

Qy 121 NKADSEQRQVREGQOQLAKEYGMDFYETSACTNLNLIKESF 162
I :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db 141 NKCDLTTKKVVDYTTAKEFADSLGIPFLETSAKNATNVEQSF 182

RESULT 12

US-09-794-257-16
; Sequence 16, Application US/09794257
; Patent No. US20020098041
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 32705, 23224, 27423, 32700, 32712, No. US2002009804A1e1
; FILE REFERENCE: 35800/209285
; CURRENT APPLICATION NUMBER: US/09/794,257
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/185,606
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Pfam accession number PF00071
US-09-794-257-16

Query Match 39.3%; Score 434.5; DB 10; Length 198;
Best Local Similarity 50.9%; Pred. No. 4.7e-36;
Matches 89; Conservative 30; Mismatches 43; Indels 13; Gaps 3;

Qy 10 RLLIGDSVGKTCLLCRFTDNEFHSHSTIGVDFKMTIEVDGKIVRIQIWDTAGQER 69
I :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db 1 KLVIGDSVGKSSLLIRFTDNKFVEEYIPTIGVDFYTKTVEVDGKTVKLQIWDTAGQER 60

Qy 70 YQITKQYRRAQGIFLVYDITSSERSYQHIMKWSDVDVEYA--PEGVQKILIGNKAD--- 124
I :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db 61 FRALRPAYRGAQGFLVYDITSRDSFENVKWLEILRHADKDNVPIVLVGNKCDLED 120

Qy 125 -----BEQKRQVREGQOQLAKEYG-MDFYETSACTNLNLIKESFTRLTELVLQ 171
I :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db 121 DEDLELTGQKRVVSTEGEALAKELGALPFMETSAKTNTNVEEAFEELAREILK 175

RESULT 13

US-09-945-173-5
; Sequence 5, Application US/09945173
; Patent No. US20020127568A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 47324, A NOVEL HUMAN G-PROTEIN AND USES
; FILE REFERENCE: 38155-20035.00
; CURRENT APPLICATION NUMBER: US/09/945,173
; CURRENT FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/229,293
; PRIOR FILING DATE: 2000-09-01

; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus amino acid sequence
US-09-945-173-5

Query Match 39.3%; Score 434.5; DB 10; Length 198;
Best Local Similarity 50.9%; Pred. No. 4.7e-36;
Matches 89; Conservative 30; Mismatches 43; Indels 13; Gaps 3;

Qy 10 RLLIGDSVGKTCLLCRFTDNEFHSHSTIGVDFKMTIEVDGKIVRIQIWDTAGQER 69
I :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db 1 KLVIGDSVGKSSLLIRFTDNKFVEEYIPTIGVDFYTKTVEVDGKTVKLQIWDTAGQER 60

Qy 70 YQITKQYRRAQGIFLVYDITSSERSYQHIMKWSDVDVEYA--PEGVQKILIGNKAD--- 124
I :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db 61 FRALRPAYRGAQGFLVYDITSRDSFENVKWLEILRHADKDNVPIVLVGNKCDLED 120

Qy 125 -----BEQKRQVREGQOQLAKEYG-MDFYETSACTNLNLIKESFTRLTELVLQ 171
I :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db 121 DEDLELTGQKRVVSTEGEALAKELGALPFMETSAKTNTNVEEAFEELAREILK 175

RESULT 14

US-09-972-529-4
; Sequence 4, Application US/09972529
; Patent No. US20020150916A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 47316, A NOVEL HUMAN G-PROTEIN AND USES
; FILE REFERENCE: 38155-20041.00
; CURRENT APPLICATION NUMBER: US/09/972,529
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: US 60/237,716
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus amino acid sequence
US-09-972-529-4

Query Match 39.3%; Score 434.5; DB 10; Length 198;
Best Local Similarity 50.9%; Pred. No. 4.7e-36;
Matches 89; Conservative 30; Mismatches 43; Indels 13; Gaps 3;

Qy 10 RLLIGDSVGKTCLLCRFTDNEFHSHSTIGVDFKMTIEVDGKIVRIQIWDTAGQER 69
I :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db 1 KLVIGDSVGKSSLLIRFTDNKFVEEYIPTIGVDFYTKTVEVDGKTVKLQIWDTAGQER 60

Qy 70 YQITKQYRRAQGIFLVYDITSSERSYQHIMKWSDVDVEYA--PEGVQKILIGNKAD--- 124
I :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db 61 FRALRPAYRGAQGFLVYDITSRDSFENVKWLEILRHADKDNVPIVLVGNKCDLED 120

Qy 125 -----BEQKRQVREGQOQLAKEYG-MDFYETSACTNLNLIKESFTRLTELVLQ 171
I :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db 121 DEDLELTGQKRVVSTEGEALAKELGALPFMETSAKTNTNVEEAFEELAREILK 175

RESULT 15

US-09-817-199A-4
; Sequence 4, Application US/09817199A
; Patent No. US20020142380A1
; GENERAL INFORMATION:

